Linkage Analysis of Molecular Markers and Quantitative Trait Loci in Populations of Inbred Backcross Lines of *Brassica napus* L.

David V. Butruille,*,† Raymond P. Guries† and Thomas C. Osborn*

*Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706 and †Department of Forest Ecology and Management, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

Backcross populations are often used to study quantitative trait loci (QTL) after they are initially discovered in balanced populations, such as F_2, BC_1, or recombinant inbreds. While the latter are more powerful for mapping marker loci, the former have the reduced background genetic variation necessary for more precise estimation of QTL effects. Many populations of inbred backcross lines (IBLs) have been developed in plant and animal systems to permit simultaneous study and dissection of quantitative genetic variation introgressed from one source to another. Such populations have a genetic structure that can be used for linkage estimation and discovery of QTL. In this study, four populations of IBLs of oilseed *Brassica napus* were developed and analyzed to map genomic regions from the donor parent (a winter-type cultivar) that affect agronomic traits in spring-type inbreds and hybrids. Restriction fragment length polymorphisms (RFLPs) identified among the IBLs were used to calculate two-point recombination fractions and LOD scores through grid searches. This information allowed the enrichment of a composite genetic map of *B. napus* with 72 new RFLP loci. The selfed and hybrid progenies of the IBLs were evaluated during two growing seasons for several agronomic traits. Both pedigree structure and map information were incorporated into the QTL analysis by using a regression approach. The number of QTL detected for each trait and the number of effective factors calculated by using biometrical methods were of similar magnitude. Populations of IBLs were shown to be valuable for both marker mapping and QTL analysis.

BACKCROSSING is widely recognized as a powerful method to study the effects of quantitative trait loci (QTL). By recurrent crossing to a single inbred genotype, one can reduce the variance caused by main and interaction effects of background QTL, allowing more precise estimates of the QTL effects under study. Backcrossing can be used after initial discovery of QTL in a balanced population, such as F_2, BC_1, recombinant inbred, or double haploid, as a means to better define the genetic position and phenotypic effects of targeted QTL. This approach has been used to study seed yield heterosis (Stuber *et al.* 1992; Graham *et al.* 1997) and domestication traits in maize (Doebley and Stec 1991; Dorweiller *et al.* 1993), and several quantitative traits in tomato (Paterson *et al.* 1988, 1990; Azanza *et al.* 1994; Alpert and Tanksley 1996; and Eshed and Zamir 1996).

Backcross lines also can be used for the initial discovery of QTL. The value of DNA markers for tagging QTL was first demonstrated using near isogenic lines of tomato developed by backcrossing with phenotypic selection (Osborn *et al.* 1987). More recently, backcrossing with molecular marker selection has been used to develop sets of introgression lines having one or a few defined segments of a donor's genome in a common genetic background (Eshed and Zamir 1994; Howell *et al.* 1996). This approach can be used to search the entire genome for donor alleles affecting a trait, but it requires a large effort to develop the lines, thorough marker coverage of the genome to avoid undetected QTL effects, and linkage information from an early generation (*e.g.*, BC_1) or another population. The more applied "advanced backcross QTL" method involves phenotypic selection while backcrossing and results in the introgression and identification of desirable donor alleles (Tanksley and Nelson 1996). However, the selection process can eliminate portions of the donor genome and create hidden pedigree structures that may lead to spurious linkage and bias in the QTL analysis. Also, molecular marker linkage information must come from an early generation (*e.g.*, BC_1) or another population.

Populations of inbred backcross lines (IBLs) developed by selfing a random set of backcross lines could be used to screen the entire genome for useful alleles affecting quantitative traits, and they provide a pedigree structure that avoids selection bias and allows for linkage analysis of molecular markers. These types of populations were first proposed by Wehrhahn and Allard (1965) to introgress desirable quantitative traits and estimate the number of genes controlling a quantitative

Corresponding author: Thomas C. Osborn, Department of Agronomy, 1575 Linden Dr., University of Wisconsin, Madison, WI 53706. E-mail: tcosborn@facstaff.wisc.edu

trait. Many populations of IBLs have been developed and evaluated phenotypically (Baker 1978; Thurling and Vijendra Das 1979; Sullivan and Bliss 1983; Rau *et al.* 1994), and populations of recombinant congenic strains, the animal counterpart to IBLs, have been used by mouse researchers (Démant and Hart 1986). However, the use of these types of populations for developing genetic linkage maps and discovering QTL has not been reported. The unequal allele frequencies inherent to such populations cause a reduction in power for detecting QTL. Despite this, populations of IBLs may be particularly useful for systematically discovering favorable alleles in unadapted germplasm and simultaneously allow for linkage analysis.

In this study, we developed and analyzed four populations of IBLs of oilseed rape (*Brassica napus* L.). *B. napus* can be classified as either spring type (sown and harvested in the same growing season) or winter type (sown in the fall and requiring vernalization before flowering in the next growing season). Genetic diversity within the species is also organized according to growth habit; molecular marker analysis indicates that there is less diversity within than between growth habits (Diers and Osborn 1994). Although winter germplasm is completely unadapted as a spring-seeded crop, it could be used to broaden the genetic base of spring-type *B. napus.* Our previous work showed that this type of introgression significantly increases the seed yield of spring hybrids (Butruille *et al.* 1999). The purpose of the present study was to learn more about the effects of introgressing winter-type germplasm into inbred and hybrid spring-type *B. napus.* We show that populations of IBLs, even with their unbalanced allele frequencies, can provide linkage information to enrich genetic maps. We also show how these populations can be used to test for associations between single marker loci and quantitative traits and to build multilocus models.

MATERIALS AND METHODS

Development of plant material: All parents used in this study were from oilseed *B. napus* cultivars having canola quality (*i.e.*, oil with <2% erucic acid and <30 μmol of aliphatic glucosinolates per gram of meal). The cultivars belong to different germplasm groups (Diers and Osborn 1994) and are adapted to different environments. The donor parent was from the German winter-type cultivar "Ceres" and the spring recurrent parents were from the spring-type cultivars "Marnoo" (Australian) and "Westar" (Canadian).

Four inbred backcross populations were developed using one or two backcrosses to the recurrent parent (Figure 1). A total of 128 BC_1S_3-BC_2S_2 pairs were planned (64 backcross pairs to Marnoo and 64 to Westar), but 5 had one of its members missing and 11 had one of its members with one selfing generation less than planned. All plants were selfed a final time to create BC_1S_4 and BC_2S_3 families for field testing. These four populations of backcross inbred lines were named "MBC_1S_3," "MBC_2S_2," "WBC_1S_3," and "WBC_2S_2." These plants also served as pollen donors to produce hybrid seed using "Topas," a European spring canola, as a "tester." Topas and
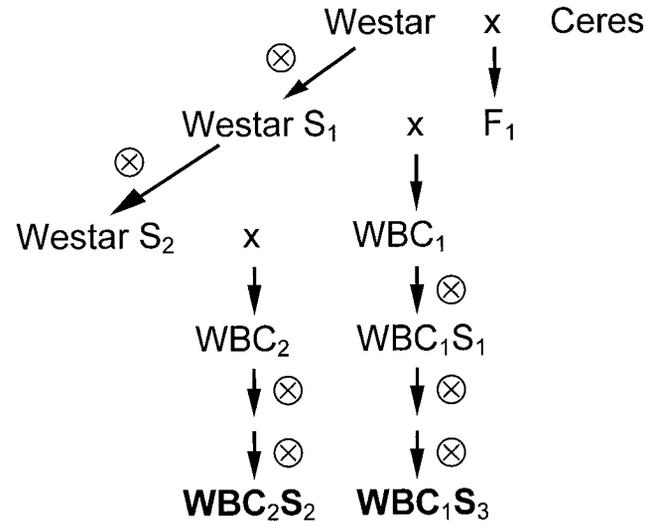
Figure 1.—Pedigree used to create two populations (WBC_1S_3 and WBC_2S_2) of inbred backcross lines (IBLs). A single plant from each cultivar was used as founder; however, a subpopulation structure was created by using two different F_1 hybrids and two different Westar S_1 plants to derive two sets of 32 BC_1S_3-BC_2S_2 pairs each. Two additional populations of IBLs (MBC_1S_3 and MBC_2S_2) were created by using the same pedigree structure, but using Marnoo instead of Westar as the recurrent parent.

its sister line "Karat" (Sernyk 1990; Diers and Osborn 1994) are known to combine well with Marnoo and Westar (Brandle and McVetty 1990; Banks and Beversdorf 1994). For each hybrid, only one female was used, and each female was hand-emasculated in the greenhouse prior to controlled pollination. The four resulting hybrid populations were named "Tp × MBC_1S_3," "Tp × MBC_2S_2," "Tp × WBC_1S_3," and "Tp × WBC_2S_2."

Trait measurements: *Yield trials:* Yield trials were conducted at the Arlington Agricultural Research Station (Columbia County, WI) during 1996 and 1997. The experiment was arranged in a split-plot design, where main plots were the eight populations (four populations of IBLs and four populations of hybrids), and subplots were individual families within each population. Subplots within main plots and main plots within year were completely randomized. Main plots were replicated three times each year (only twice for IBLs of WBC_2S_2 and MBC_2S_2 in 1997, due to insufficient seeds). Several checks were also used, including the recurrent parents (Westar and Marnoo), the tester (Topas), other commercial open-pollinated and hybrid cultivars, and all possible F_1s among Westar, Marnoo, Topas, and Ceres. Ceres was not included because it would fail to flower under our conditions. Standard field practices were used, including incorporation of herbicide (Trifluralin at 2 liters/ha) and fertilizer (100 kg of N/ha) prior to planting and additional hand weeding when necessary. No other pesticide application was needed.

In 1996, each subplot consisted of two rows, 3 m long and 0.30 m apart and separated from the next plot by a guard row (at 0.30 m) sown with bulked seed from the same population. The seeding rate was 110 seeds/m². In 1997, each subplot was seven rows wide (0.15 m between rows) and 2.40 m long, adjacent subplots were 0.30 m apart and the seeding rate was 87 seeds/m². For each plot, the date when half of the plants had at least one open flower was recorded. Before harvest in 1996, we measured plant height in each plot. When seeds started to turn color, the plots were hand harvested, dried, and threshed. The center 1.8 m of all plot rows (two rows in

1996, seven in 1997) was harvested. We measured the seed yield and 1000-seed weight for each plot. The oil content of the inbred families was determined by David Syme, PGS Canada, from a composite sample of all the replicates of each entry within each year using the NMR method in which percentage oil is calculated by weight on a whole seed basis at 0% moisture.

Trait analysis: Statistical analyses were conducted using the MIXED procedure of SAS (Littell *et al.* 1996). Year, recurrent parent, backcross level, and hybridity were treated as fixed effects while entries within population and blocks within year and population were treated as random effects to study differences at the population level. In 1997, an unplanned systematic pattern appeared in the field due to staggering of subplots at planting, combined with lodging and white mold infection (*Sclerotinia sclerotiorum*). This effect was included in the model by using an indicator variable to describe the position of every subplot in the field relative to other subplots. Broad sense heritability estimates were obtained for each population and each trait by using the estimates of the variance component parameters calculated with the MIXED procedure as

$$\hat{h}^2 = \hat{\sigma}^2_{entry}/(\hat{\sigma}^2_{residual} + \hat{\sigma}^2_{entry}).$$

Data were also analyzed considering individual entries within the population as fixed effects to derive least-squares estimates for the biometrical and QTL analyses. Subpopulation, as defined by which F_1 and S_1 within each pedigree were used in the backcrossing steps (legend of Figure 1), was included as an effect in the linear model prior to biometrical analysis. Least-squares estimates for the checks were obtained for the biometrical analysis.

IBLs were originally proposed to estimate the minimum number of genes controlling a trait (Wehrhahn and Allard 1965). To calculate this number, we determined what was the proportion of lines (or hybrids) significantly different from the parents (or parental hybrids; Mulitze and Baker 1985). Those were lines (or hybrids) with a mean value outside of the confidence interval,

$$[\hat{\mu}_{parent} - 1.96\sqrt{\hat{\sigma}^2_{error\ parent} + \hat{\sigma}^2_{error\ IBL}},$$

$$\hat{\mu}_{parent} + 1.96\sqrt{\hat{\sigma}^2_{error\ parent} + \hat{\sigma}^2_{error\ IBL}}].$$

The value of $\hat{\sigma}^2_{error\ parent}$ was obtained for each parent (or parent $\times$ Topas), and $\hat{\sigma}^2_{error\ IBL}$ was obtained for each IBL (or IBL $\times$ Topas) using the MIXED procedure of SAS. These error variances generally correspond to $MSE_{subplot}/n_{replicates}$ but depart from this value for entries with missing data. From this separation of the population, we calculated the observed proportion of nonparental lines, $\hat{d}$, and the estimated number of loci, $\hat{k}$, as

$$\hat{k} = \ln(1 - \hat{d})/\ln\left[1 - \frac{1}{2}^{b+1}\right],$$

where b is the number of backcrosses, and the 95% confidence interval for $\hat{k}$ is

$$\left(\ln[1 - \hat{d} + 1.96\sqrt{\hat{d}(1 - \hat{d})/m}]/\ln\left[1 - \frac{1}{2}^{b+1}\right],\right.$$

$$\left.\ln[1 - \hat{d} - 1.96\sqrt{\hat{d}(1 - \hat{d})/m}]/\ln\left[1 - \frac{1}{2}^{b+1}\right]\right),$$

where m is the sample size.

Two types of analyses of covariance were conducted using the least-squares estimates of the means of the traits studied. The first one tried to assess the impact of the paired structure (BC_1S_3-BC_2S_2) in the pedigree. Means of the BC_2S_2 lines were used as covariates in a model where the means of the paired

BC_1S_3 lines were the dependent variables. This was also done for hybrids. Other effects in this model were parents and subpopulation within parents. In the second type of analysis of covariance, performed to detect possible pleiotropic effects, yield was the dependent variable and the other traits were used as covariates. It was done within the IBL and within the IBL test-crossed onto Topas. Inbred yield was also used as a covariate for hybrid yield.

Molecular markers: The DNAs for restriction fragment length polymorphism (RFLP) genotyping of the inbred lines, parents, original F_1s, and tester were extracted from bulked tissue of 24 seedlings from each entry. We otherwise proceeded as described in Ferreira *et al.* (1994). Most probes (genomic and cDNA probes) were previously used by Ferreira *et al.* (1994) and by Thormann *et al.* (1996). Two *Arabidopsis thaliana* genomic probes, mi193 and mi259, were also used (Liu *et al.* 1996; obtained from the Arabidopsis Biological Resource Center, Ohio State University). Restriction endonucleases used were *Eco*RI and *Hin*dIII. Because *B. napus* is an amphidiploid derived from two diploid species (*B. rapa* and *B. olearacea*), which themselves have some degree of genome duplication, the polymorphisms detected did not always correspond to the same loci mapped previously. Loci in the present populations were considered identical to the one previously mapped when RFLP alleles had identical molecular weights. RFLP alleles at some loci behaved as dominant markers, but most were codominant. We scored both types of loci using the following system ("*D*" represents the allele from the donor parent and "*R*" the recurrent parent allele):

$C = \{DD\}$ — homozygous codominant or recessive donor allele;

W (or M) $= \{RR\}$ — homozygous codominant or recessive recurrent allele;

$H = \{DR\}$ — heterozygous codominant alleles;

$B = \{DD, DR\}$ — homozygous or heterozygous dominant donor allele;

$V = \{DR, RR\}$ — homozygous or heterozygous dominant recurrent allele;

$X = \{DD, DR, RR\}$ — no information.

Autoradiographs were scored at least three times to reduce possibilities of misscoring. For some loci, the original parent was heterozygous (1 of 122 loci for Westar and 18 of 122 loci for Marnoo), and phase had to be assessed in a recursive fashion by first mapping the locus in both phases and then choosing the phase that resulted in the most likely configuration [reduced recombination and highest logarithm of the odds (LOD) score].

Linkage analysis: To our knowledge, none of the genetic mapping software commonly used would perform multipoint analysis and mapping on this pedigree. However, several mapping procedures accept two-point information to build genetic maps (two-point information consists of an estimated recombination frequency and a corresponding LOD score). Thus, we chose to perform the exact two-point analysis and use the resulting information to enrich a previous map built in a population of 105 F_1-derived doubled-haploid lines (DH lines; Ferreira *et al.* 1994). Several programs with which to build composite maps are available, including MAP+ (Collins *et al.* 1996), a software package developed to build composite genetic maps based on multiple pairwise information. This software requires that an approximate order of loci be known and uses all the pairwise information that can be obtained. MAP+ allowed us to test for heterogeneity of the recombination fraction between populations, and we used this to identify regions where merging was likely to give erroneous marker orders. Because this software requires a reasonably accurate starting order, we added new loci sequentially, beginning with

the loci that had been mapped previously in a population of DH lines (Ferreira *et al.* 1994; Thormann *et al.* 1996; Osborn *et al.* 1997) using MAPMAKER (Lander *et al.* 1987) and data for 200 RFLP and 260 amplified fragment length polymorphisms (AFLP; Vos *et al.* 1995) loci. For this study, we used all of the RFLP loci plus 10 AFLP loci that gave expanded genome coverage, and we determined the pairwise recombination values for the population of DH lines by simply counting the recombinant (r) and nonrecombinant gametes (s) and then calculating the recombination frequency as $\theta = r/(r + s)$ and the LOD score as

$$\text{LOD} = (r + s)\,[\log(2) + \theta \log(\theta)$$
$$+ (1 - \theta)\log(1 - \theta)\,] \quad \text{if } r > 0$$

and

$$\text{LOD} = s[\log(2)\,] \quad \text{if } r = 0.$$

For the populations of IBLs, we built tables of joint-probability distributions of the BC_1S_3 and the BC_2S_2 members of each pair. The joint probability distribution formulas are given as a function of the recombination fraction in a BC_1S_3-BC_2S_2 pair when both loci were homozygous in the parents (appendix a, *A1*). Failure of one of the loci to be homozygous in the parents or one of the lines to be selfed one time less required construction of additional tables (appendix a, *A2*). To reduce computing time, "grid searches" were conducted using successively more refined steps: in the first step we assumed that all loci were homozygous in the parents and that all lines were at the intended selfing level. The table was constructed at 1% increments (selection of the recombination fraction that gave the highest LOD score after scanning in 1% steps from 0 to 50%). The second step, still at 1% increments, was used to recalculate the values for loci that were heterozygous in the recurrent parent and corrected for the actual selfing level of each line. The third step served to refine the precision of the table from 1 to 0.1% (appendix b).

To compare the informativeness of each population (the DH lines, the WBC_1S_3/WBC_2S_2 IBLs, and MBC_1S_3/MBC_2S_2 IBLs), we used the recombination frequency and the LOD score of each pair of loci belonging to the same linkage group and estimated the equivalent number of fully informative F_1 gametes (n_{eq}) needed to obtain the same information:

$$n_{eq} = \text{LOD}/[\log(2) + \theta\log(\theta)$$
$$+ (1 - \theta)\log(1 - \theta)\,] \quad \text{when } 0 < \theta < 0.5;$$
$$n_{eq} = \text{LOD}/\log(2) \quad \text{when } \theta = 0.$$

The values obtained were then averaged within intervals of 2% recombination (*i.e.*, 0–2%; 2–4%; . . . ; 48–50%) so that the loss of information in the IBLs due to additional meioses could be documented. The use of the actual data, instead of expectations, allowed us to examine the overall quality of the information (considering missing data, dominant markers, heterozygous parents, etc.). The final map was drawn assuming an interference parameter of 0.5, for which the Rao mapping function is identical to the Kosambi mapping function.

QTL analysis: *Assigning genotypic probabilities at loci:* Once a composite genetic map was in hand (*i.e.*, locus orders and distances), an additional computer program was written to assign to every locus in each individual a probability of being either *RR*, *DR*, or *DD* [$P(RR)$, $P(DR)$, or $P(DD)$, respectively, as described in appendix c]. If complete marker information was available at this locus for this individual, the probability would be 1 that it is the genotype scored and 0 for the remaining two. Otherwise, these probabilities were calculated using the partial information available for this locus (*e.g.*, dominant allele present) in this individual and in the other

member of the backcross pair, as well as the information available from up to three partially or completely informative flanking loci (within a distance of 43 cM, which corresponds to 35% recombination with the Kosambi mapping function) on each side of this locus in both members of the pair. When local marker information was not available the resulting probabilities were calculated on the basis of the pedigree structure alone. Probabilities at positions within intervals can be calculated as for loci with missing data. As a simplifying assumption, when a locus was heterozygous in the recurrent parent, we assumed that a QTL at or near it would still be homozygous. For this reason, the IBL marker data at this locus were recoded (*M* or *W* remained as such, but *H* was converted to *V*, and *C* and *B* were both converted to *X*) prior to calculating the genotypic probabilities of a coincident QTL.

Backward elimination of main effects: The least-squares estimates of each quantitative trait were used as dependent variables in a linear model that was simplified by backward elimination. The full model included "parent" (Westar or Marnoo), "backcross" (one or two), "parent by backcross" interaction, "subpopulation" (within parent), "subpopulation by backcross" (within parent) interaction, and "translocation." This last parameter was included to account for the effects of a reciprocal translocation present in Marnoo and Westar with respect to Ceres (our unpublished data). Segregating progeny from a cross between Marnoo or Westar and Ceres could lose either the *B. rapa* or the *B. oleracea* homeologous fragment involved in the translocation, resulting in five states for the translocation in this model: 0, 1, 2, 3, and 4 doses of the *B. rapa* homeolog, and simultaneously 4, 3, 2, 1, and 0 doses of the *B. oleracea* homeolog. Five RFLP probes (WG5A1, EC3E12, WG7F5, TG5D9, and WG2A3) that hybridize to loci on the translocation were used to determine the status of each IBL with respect to this translocation.

The pairing of IBL between the two populations due to a common BC_1 ancestor was not considered in this model. Although this might have resulted in a slight overestimation of the effects of QTL, the possibility that a marker-trait association is the result of the marker indicating a good family instead of indicating physical linkage is much smaller than with advanced-backcross QTL when all advanced lines are derived from a few BC_1 or BC_2 ancestors. To obtain more information on the potential bias, an analysis of covariance was done for each trait using the value of the BC_1S_3 line as a dependent variable and the value of the associated BC_2S_2 lines as a covariate.

Forward selection of marker loci: Marker loci were added to the reduced model using forward selection. For the inbred populations the marker information was divided into additive (a) and dominance (d) components:

$$P_a = P(DD) - P(RR); \quad P_d = P(DR).$$

For traits measured in the hybrid populations only an additive component could be estimated. In this case, $\bar{P}_a$ actually represents $P(DT) - P(RT)$, where *T* indicates the tester allele. When a locus is heterozygous in the IBL parent (*DR*), half of its testcrossed progeny is *RT* and half is *DT*.

At each cycle of testing the most significant marker locus was added to the multilocus model until no other loci with a significant effect could be added. Loci that were segregating only in the DH population were not included in the model. Loci that segregated only in the Westar or in the Marnoo background were included if flanking loci segregated in the other background; otherwise, these loci were included when each background was analyzed separately. The addition of each locus required the testing of several hypotheses to choose the best model, such as the presence of an average additive effect, an average dominance effect, and a difference of effects

TABLE 1

Heritabilities ($\hat{h}^2$) and estimated number of genes ($\hat{k}$) for agronomic and seed traits in inbred and hybrid populations of inbred backcross lines

IBLs	Inbred populations			Hybrid populations		
	$\hat{h}^2$	$\hat{k}$	95% C.I.[a] for $\hat{k}$	$\hat{h}^2$	$\hat{k}$	95% C.I. for $\hat{k}$
Flowering time						
MBC$_1$S$_3$	0.94	5.3	(4.0–7.4)	0.78	3.0	(2.1–4.2)
MBC$_2$S$_2$	0.91	5.7	(3.9–7.9)	0.74	2.3	(1.3–3.5)
WBC$_1$S$_3$	0.96	7.5	(5.6–11.6)	0.86	5.3	(3.9–7.5)
WBC$_2$S$_2$	0.91	10.7	(8.0–15.1)	0.74	4.8	(3.3–6.9)
Plant height						
MBC$_1$S$_3$	0.40	1.1	(0.7–1.7)	0.14	1.4	(0.8–2.0)
MBC$_2$S$_2$	0.32	0.9	(0.3–1.5)	0.18	0.6	(0.1–1.2)
WBC$_1$S$_3$	0.50	2.4	(1.6–3.4)	0.18	0.7	(0.3–1.2)
WBC$_2$S$_2$	0.34	3.5	(2.2–5.2)	0.22	1.6	(0.7–2.5)
Seed weight						
MBC$_1$S$_3$	0.64	3.0	(2.1–4.2)	0.44	2.6	(1.8–3.7)
MBC$_2$S$_2$	0.69	5.4	(3.7–7.6)	0.51	1.0	(0.3–1.7)
WBC$_1$S$_3$	0.59	2.8	(1.9–3.9)	0.34	1.0	(0.5–1.6)
WBC$_2$S$_2$	0.46	3.1	(1.9–4.6)	0.37	1.4	(0.6–2.3)
Oil content						
MBC$_1$S$_3$	0.76	1.7	(1.1–2.5)	—	—	—
MBC$_2$S$_2$	0.74	1.7	(0.8–2.7)	—	—	—
WBC$_1$S$_3$	0.69	3.0	(2.1–4.3)	—	—	—
WBC$_2$S$_2$	0.51	2.7	(1.6–4.0)	—	—	—
Seed yield						
MBC$_1$S$_3$	0.57	2.5	(1.7–3.5)	0.10	0.3	(0.1–0.6)
MBC$_2$S$_2$	0.43	8.0	(5.8–11.1)	0.05	0.5	(0.0–1.0)
WBC$_1$S$_3$	0.50	1.9	(1.2–2.7)	0.18	2.2	(1.5–3.2)
WBC$_2$S$_2$	0.47	2.9	(1.7–4.2)	0.11	1.0	(0.3–1.8)

[a] C.I., confidence interval.

between the Westar- and the Marnoo-derived populations. A significance threshold of $P < 0.002$ (*i.e.*, LOD of 2.7) was chosen for adding a new locus into the model. The percentage of the variance explained by a given multilocus model was calculated as

Variance explained

$$= 100 \times (SS_{\text{main effects + all loci selected}} - SS_{\text{main effects}})/SS_{\text{Total}},$$

where SS is the sum of squares obtained from the analysis of variance table. The formula to calculate the additional variance explained by incorporating a new locus in the model was

Var. expl. $= 100$

$$\times (SS_{\text{main effects + all loci selected}}$$

$$- SS_{\text{main effects + all loci selected but the last one}})/SS_{\text{Total}}.$$

Estimating the effects of allele substitution on a trait: We equated the expected mean performance of the selfed progeny of an IBL to

$$\mu + a\, P_a + d\, P_d.$$

The effect of substituting *RR* by *DD* at the locus under consideration was thus estimated to be $2a$, and the actual dominance deviation was twice the value estimated by d because only half the selfed progeny of a heterozygous IBL was expected to be heterozygous.

The expected mean performance of the testcrossed progeny of an IBL can then be written as

$$\mu + a[P(DT) - P(RT)].$$

The value $2a$ then becomes the estimated effect of substituting *RT* by *DT* in the hybrid.

RESULTS AND DISCUSSION

Trait analysis: For each trait, inbred populations had higher heritabilities than hybrid populations and BC$_1$ populations had higher heritabilities than BC$_2$ populations (Table 1). The higher heritabilities were associated with higher variances among lines within these populations. This expected trend also can be seen by the distribution of least-squares estimates of means of IBL or hybrids for each trait within each population (Figures 2–6). The estimated number of genes seems to be related more to the heritability of the trait than to what one might assume *a priori* about its underlying genetic control. For example, we might expect flowering time to be controlled by a smaller number of loci than seed yield; however, the inbred-backcross method led us to conclude that many more loci influence it (Table 1). These results may be due to the fact that this method
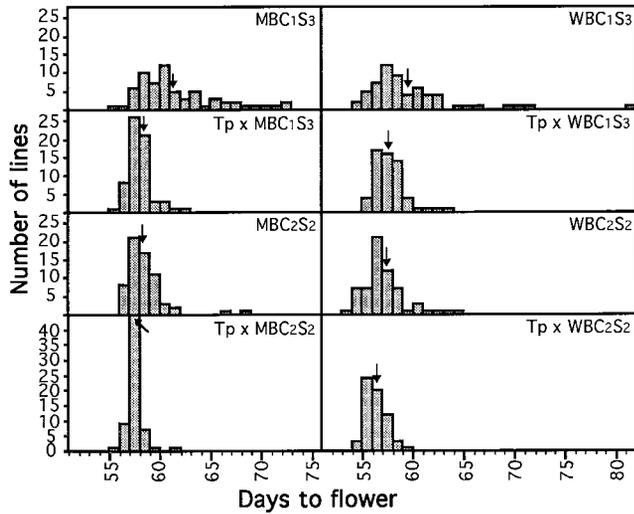
Figure 2.—Histograms of least-squares estimates of the means for days to flowering in each population. The arrows indicate the population means. Estimates of the parental means and other checks (in days after planting) are Marnoo, 58.0; Marnoo × Ceres, 68.5; Westar, 55.3; Westar × Ceres, 66.3; Topas, 56.6; Topas × Marnoo, 57.2; Topas × Westar, 55.9; and Topas × Ceres, 67.0.

can only detect loci where allele substitution leads to changes in phenotype $> 1.96\sqrt{\hat{\sigma}^2_{\text{error parent}} + \hat{\sigma}^2_{\text{error IBL}}}$, a result that requires few alleles of large phenotypic effects or very high heritabilities, both unlikely for more complex and elusive traits such as yield. Furthermore, as Baker (1978) pointed out, a line that differs significantly from the recurrent parent could also result from
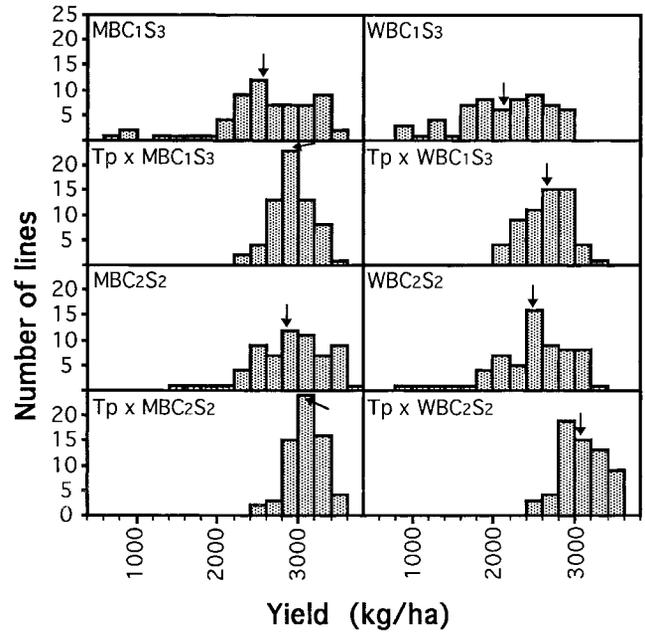
Figure 4.—Histograms of least-squares estimates of the means for seed yield in each population. The arrows indicate the population means. Estimates of the parental means and other checks (in kilograms per hectare) are Marnoo, 2241; Marnoo × Ceres, 3508; Westar, 2426; Westar × Ceres, 3810; Topas, 2637; Topas × Marnoo, 3014; Topas × Westar, 3098; and Topas × Ceres, 3249.

the accumulation of several minor factors under a polygenic model.

Main effects [year, recurrent parent (Marnoo or Westar), backcross level (one or two backcrosses), and
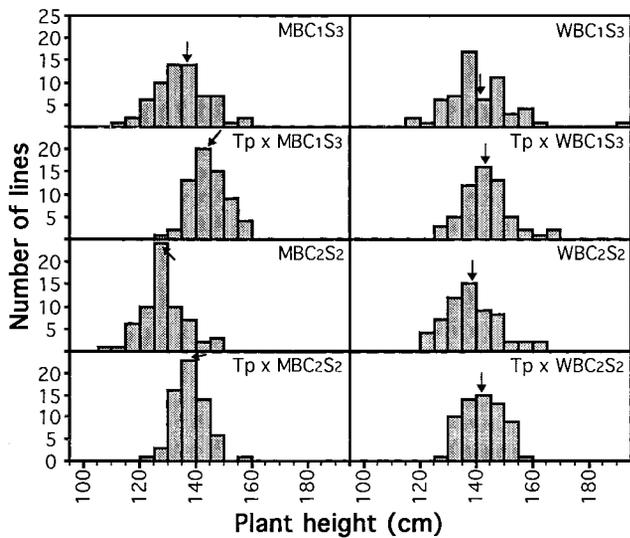
Figure 3.—Histograms of least-squares estimates of the means for plant height in each population. The arrows indicate the population means. Estimates of the parental means and other checks (in centimeters) are Marnoo, 127; Marnoo × Ceres, 166; Westar, 126; Westar × Ceres, 164; Topas, 133; Topas × Marnoo, 135; Topas × Westar, 136; and Topas × Ceres, 185.
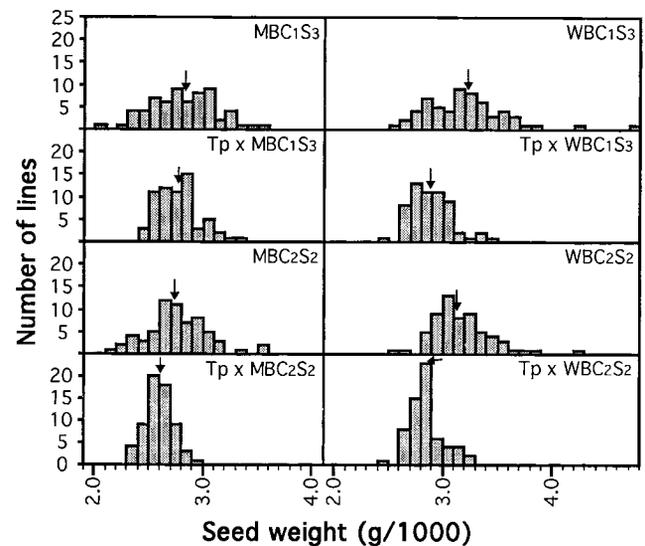
Figure 5.—Histograms of least-squares estimates of the means for seed weight in each population. The arrows indicate the population means. Estimates of the parental means and other checks (in g/1000 seeds) are Marnoo, 2.532; Marnoo × Ceres, 2.779; Westar, 3.210; Westar × Ceres, 3.307; Topas, 2.466; Topas × Marnoo, 2.533; Topas × Westar, 2.815; and Topas × Ceres, 2.940.
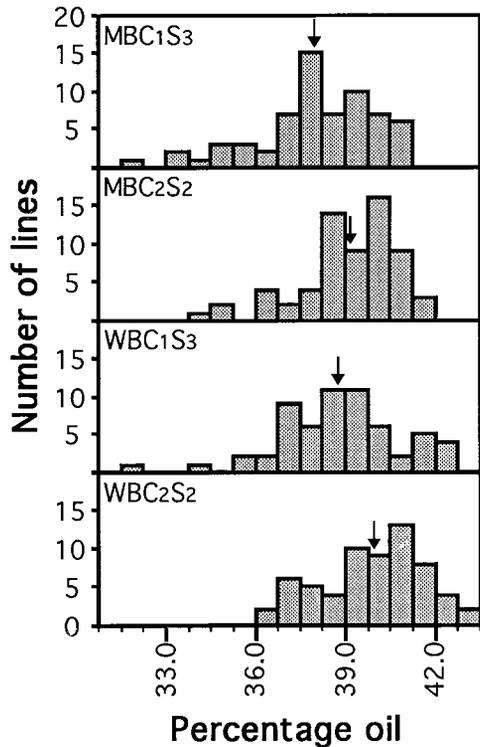
Figure 6.—Histograms of least-squares estimates of the means for oil content in each population. The arrows indicate the population means. Estimates of the parental means (in percentage oil) are Marnoo, 39.7 and Westar, 40.9.

hybridity (inbred or hybrid populations)] were significant for yield, seed weight, percentage oil, days to flowering, and plant height (Table 2). The yield of hybrid populations exceeded the yield of inbred populations both years (on average by 17%). Yield, seed weight, and oil content were greater the first year due to better growing conditions. Populations for which Westar was the recurrent parent flowered earlier, had larger seeds, higher oil content, and were taller. Some of the two-way interactions (mostly the parent-by-hybridity interaction) were also significant. In general, the three- and four-way interactions were not significant. At the subplot level, entries and systematic pattern in the field in 1997 were significant. Subpopulation was significant except for seed yield in 1996 (Table 2).

When the means of BC_2S_2 lines, or hybrids, were used as a covariate in a model where the means of the paired BC_1S_3 lines were the dependent variables, they significantly ($P < 0.05$) reduced the error variance for days to flowering, seed weight, plant height, and oil content for the inbred populations (and seed weight for the hybrid populations). However, the maximum average correlation between pairs, adjusted for the design, was only 0.32 (seed weight in the hybrids). Thus, at most 10% (100×0.32^2) of the variance of the BC_1S_3 lines or hybrids could be attributed to its linear regression on the means of the BC_2S_2 lines or hybrids. Correlations in the Westar background were not significantly different

from the correlations estimated in the Marnoo background. The QTL analysis was simplified by ignoring this pairing of lines.

Significant associations were also found between yield and other agronomic traits. For the inbreds, days to flowering had a significant negative association with yield ($P < 0.0001$; -42 kg/ha/day; $r = -0.27$), an association due mainly to a few low yielding, very late flowering lines (Figure 2). Oil content of the inbreds was significantly associated with yield ($P < 0.0001$; $+200$ kg/ha/%; $r = 0.69$). Seed weight was positively associated with seed yield only in the Marnoo background ($P < 0.01$; 480 kg/ha per g/1000; $r = 0.26$). For the hybrids, days to flowering had a significant positive association with yield only in the Westar background ($P < 0.001$; 67 kg/ha/day; $r = 0.31$) and inbred yield had a significant positive association with hybrid yield in both backgrounds ($P < 0.0001$; $+0.161$ kg/kg; $r = 0.33$).

Linkage analysis: Molecular marker information from the populations of IBLs was used to enrich, with 72 new RFLP loci, a previously published genetic map of *B. napus* based on a population of DH lines (Ferreira *et al.* 1994; Thormann *et al.* 1996; Osborn *et al.* 1997). Loci were added to a linkage group if they had a recombination fraction <30% and a LOD score >4 with at least one member of that linkage group. At short distances, information from the populations of IBLs contributed substantially to the genetic map. However, for recombination fractions >20%, the combined information of 251 IBLs was not as useful as the 105 F_1-derived DH lines. And often only one of the genetic backgrounds (Marnoo or Westar) provided information on linkage between two loci, in which case 50 F_1 gametes would be more informative (Figure 7). This was expected because additional meioses lead to more recombination between adjacent marker loci. This observation justifies a mapping strategy that would first place loci using information from the population of DH lines and then use MAP+ to insert new loci mapped in the IBLs only. However, with molecular markers increasingly more available, several hundreds of closely spaced marker loci can be used to build maps. This would reduce greatly the need for early balanced generations to connect more distant loci, thus avoiding the need to merge maps. The composite map has 276 loci in 19 linkage groups, two triplets, three pairs, and 4 unlinked loci (Figure 8).

The reciprocal translocation between homeologous segments of N7 and N16 (Sharpe *et al.* 1995) was present in both recurrent parents and resulted in abnormal segregation in the IBLs. No other loci from N16 segregated in this population (unless it was one of the unlinked loci listed in Figure 8). Loci contained in this translocation (scored with probes WG7F5, WG2A3, TG5B9, WG5A1, and EC3E12) clustered and were used to count the number of copies of each homeologous fragment. To do so, we identified which RFLPs were

TABLE 2

Significance of effects in the analysis of the field trial for yield and other traits

Effects	Days to flowering	Plant height (cm)	Yield (kg/ha) 1996–1997	1996	1997	Seed weight (g/1000)	% oil
Year	****	—	****	—	—	****	****
Parent	****	**	****	NS	****	****	****
Backcross level	****	**	****	**	****	*	****
Hybridity	****	***	****	***	****	****	—
Y × P[a]	*	—	***	—	—	NS	****
Y × B	NS	—	NS	—	—	**	NS
Y × H	*	—	*	—	—	**	—
P × B	NS	NS	NS	*	NS	NS	NS
P × H	**	*	**	NS	***	***	—
B × H	****	NS	NS	NS	NS	NS	—
Y × P × B	NS	—	*	—	—	NS	*
Y × P × H	NS	—	*	—	—	*	—
Y × B × H	NS	—	NS	—	—	NS	—
P × B × H	*	NS	NS	NS	NS	NS	—
Y × P × B × H	NS	—	NS	—	—	NS	—
Entry (P × B × H × S)	****	****	****	****	****	****	****
Subpopulation (P)	****	**	****	NS	****	****	*
In/out (Y)[b]	**	—	****	—	****	****	—
N/S (Y)[b]	NS	—	****	—	****	****	—
SD[c]	0.83	9.9	495	506	431	0.235	1.14
SE	0.34	5.7	202	292	249	0.096	0.80

* Significance level of 5%; **significance level of 1%; ***significance of level of 0.1%; ****significance level of 0.01%.

NS, not significant; —, effect does not apply.

[a] Y, P, B, H, and S are abbreviations used in the interaction terms for year, parent, backcross level, hybridity, and subpopulation, respectively.

[b] In/out and N/S (north/south) refer to the staggered position of subplots in 1997.

[c] Calculated as the square root of the subplot mean square error plus block variance.

alleles on the *B. rapa* homeolog (N7) and which RFLPs were alleles on the *B. oleracea* homeolog (N16) by screening Southern blots containing DNA from the three species (*B. napus*, *B. oleracea*, and *B. rapa*) and comparing the molecular weight of RFLPs among species. We assumed that the total number of alleles from the two homeologs combined was always four. The effect of this translocation on phenotypic traits was determined by testing whether differences in the number of *B. rapa* alleles were significantly associated with the expression of the trait.

At other loci, heterogeneity between recombination fractions in the different populations was significant for marker locus *wg1g5c* only. As we could not identify whether it was due to a translocation or other abnormality, linkage group N1 was split into N1 proper and N1T, which contains only new marker loci and *wg1g5c.* Linkage was detected between loci on N1 and N11 and these groups were split using a more stringent criterion, which required a minimum LOD score of four and maximum recombination of 20% (likewise for N10 and N19). A simpler population structure is required to elucidate these abnormalities. Sharpe *et al.* (1995) already re-

Figure 7.—Equivalent number of fully informative F_1 gametes (n_{eq}), which would give pairwise information (percentage recombination and LOD score) equal to that observed for the population of 105 DH lines and the IBL populations (with 124 and 119 IBLs in the Marnoo and Westar backgrounds, respectively). Values are averaged for intervals of 2% recombination.

Figure 8.—Composite map of *Brassica napus.* Vertical lines are linkage groups, and all designations to the right are locus symbols but only those loci that segregated in the Westar-derived IBLs are italicized, only those loci that segregated in the Marnoo-derived IBLs are underlined, and only those loci that segregated in the DH population are in bold. Distances between loci are indicated by the scale shown in centimorgans, using the Kosambi mapping function. Linkage group designations follow the convention of Parkin *et al.* (1995) based on the integration of that map with the map of Ferreira *et al.* (1994) by Osborn *et al.* (1997).

ported chromosomal aberrations involving N1 and N11. Other translocations discussed by Sharpe *et al.* (1995) were not conspicuous in this population.

Several areas of the composite map were poorly marked in the IBLs because some loci mapped previously lacked polymorphisms between donor and recurrent parents or were too difficult to score reliably. The top of N2 and the bottom of N14 lack coverage in both IBL populations, and an alternative order of loci is equally likely in N4 and N18 due to the lack of sufficient loci common to the IBL and DH populations (only *wg6f10* in N4) or between the IBL populations (in N18 *wg1e3c* has equal probability of being between *tg5b2a* and *wg3g9a*). The order chosen for *ec2e4*, *wg6f10*, and *wg6a12* on N4 reflects the order of loci as mapped previously in the *B. rapa* homolog (linkage group 10; Teutonico and Osborn 1994). The IBL populations that had Marnoo as a recurrent parent also were poorly covered at the top of N5, N17, and N19. Genomes of

the Westar-derived IBLs were covered more thoroughly by markers.

QTL analysis: *Flowering time:* The QTL analyses for flowering time essentially confirm results reported by Ferreira *et al.* (1995) and Osborn *et al.* (1997) from analysis of a DH population having different spring and winter parents. Seven loci having significant effects upon flowering time of the IBLs could be added into a multilocus model, with marker locus *wg6b10a* (on N2) showing the most significant effect. This locus was identified as closely linked to *VFN1*, the major flowering time QTL identified previously, while *ec3f12b* had a significant effect in the region where *VFN3* was postulated (Ferreira *et al.* 1995; Osborn *et al.* 1997). Osborn *et al.* (1997) proposed the presence of *VFN2* in a region of N10 marked by *ec3g3c*; although an effect in the same direction was found at this locus, it was not significant in this experiment and this locus was not included in the multilocus model. However, several new regions

Figure 8.—*Continued.*

were identified as contributing to this trait. The most important was in linkage group N12 at marker locus *ec2d1b.* There was also a significant digenic epistasis between *ec2d1b* and *wg6b10a* in the Westar background (Table 3).

Satagopan *et al.* (1996) suggested that a second flowering QTL, named *FN1* by Osborn *et al.* (1997), could be present on N2, the same linkage group as *VFN1.* This conclusion was based upon analyzing flowering data from DH lines subjected to 8 wk of vernalization (Ferreira *et al.* 1995). Our study lends credence to this conclusion, as we find that *ec3g3b* (on N2) has a significant effect after *wg6b10a* has been included in the model (Table 3). The final multilocus model explained 58.8% of the variance after backcrossing level was included in the model.

The main flowering time QTL were also detected in the populations of IBLs testcrossed onto Topas. The final multilocus model contained *wg6b10a* (N2), *ec2d1b*

(N12), and *ec3f12b* (N3). When the two backgrounds were analyzed separately, the unlinked marker locus *tg5b2* contributed significantly to the model in the Marnoo background and locus *wg6e1* on linkage group N17 in the Westar background. The inclusion of marker locus information in a model already containing parent, backcross, and the interaction between these, explained an additional 31.4% of the variance estimated in these hybrid populations. The magnitude of the effects estimated in the hybrids is less than half the magnitudes of the effects in the inbreds for *wg6b10a, ec3g3b, ec3f12b,* and *ec2d1b,* which mark significant QTL in both types of populations (Table 3). This result supports the conclusion that early flowering alleles are partially dominant over late flowering alleles if we assume that alleles at flowering loci in Topas resemble alleles in Marnoo or Westar.

Plant height: QTL analysis for plant height of the IBLs and hybrids pointed essentially toward the same chro-

TABLE 3

Marker loci significantly associated with flowering time (days after planting) and plant height (cm) and effect on the traits of an allelic substitution at these loci

Marker locus	Linkage group[a]	IBL			IBL × Topas		
		Sign.	Variance explained[b] (%)	Effect of $RR \to DD$[c] (days or cm)	Sign.	Variance explained[b] (%)	Effect of $TR \to TD$[c] (days or cm)
Flowering time							
wg6b10a	N2	****	22.6	6.2	****	12.9	1.4
ec3g3b	N2	****	1.8	1.6	**	1.2	0.4
ec3f12b	N3	****	4.0	2.2	****	3.8	0.8
ec2e4b	N7	***	1.5	1.3	NS	—	—
wg6g9a	N8	***	1.9	[1.6, −0.9]	NS	—	—
wg3h8a	N9	****	2.3	[−1.8, 0.8]	NS	—	—
ec2d1b	N12	****	26.0	[2.7, 8.3]	****	12.3	[0.7, 1.8]
wg6e1	N15	NS	—	—	***	2.7	[NS, 0.8]
tg5b2d	un.	NS	—	—	****	2.5	[−3.8, NS]
wg6b10a × ec2d1b[d]		****	3.8	[NS, −3.3]	NS	—	—
Plant height							
wg6b10a	N2	****	11.60	10.6	****	8.5	5.8
ec4d11b	N3	****	4.20	6.4	NS	—	—
ec2d1b	N12	****	11.30	[5.8, 17.7]	****	5.2	4.8
ec3g3a	N13	***	2.70	5.0	NS	—	—

** Significance level of 1%; ***significance level of 0.1%; ****significance level of 0.01%. NS, not significant; Sign., significance level of marker locus; —, effect does not apply.

[a] Nomenclature of Parkin *et al.* (1995)

[b] Percentage of the variance explained by a locus when added last to the model.

[c] *R*, *D*, and *T* indicate the recurrent parent, donor parent, and Topas alleles at the locus, respectively. $RR \to DD$ represents the substitution of the *RR* genotype by the *DD* genotype (more details are provided in materials and methods). When effects differed significantly in each background ($P < 0.01$), estimates of allele substitution are reported in brackets for each background separately, [Estimate_Marnoo background, Estimate_Westar background]. Otherwise, a pooled estimate is given.

[d] Estimates the interaction as difference between the effect of the joint substitution of alleles at both loci and the sum of the effect of allele substitution at each locus separately.

mosomal regions as for flowering time (Table 3). This is probably due to a pleiotropic effect of the flowering time genes on plant height, rather than linkage between QTL affecting these two traits, because later lines grew taller by producing more leaves before the apical meristem was converted into an inflorescence meristem. The inclusion of marker loci in the linear model accounted for 27.7% of the variance in the IBLs and 13.3% in the hybrids.

Seed yield: The effect of the translocation between N7 and N16 had a much greater impact on seed yield of the IBLs than on the hybrids: with 4 d.f. it explained an additional 17.4% of the variance when added to the linear model after parent and backcross effects ($P < 0.0001$), compared to only 4.2% ($P < 0.01$) in the hybrids. The absence of the chromosome segment homologous to *B. rapa* had the most deleterious effect, and the balanced condition was the most desirable configuration. Adding the information from marker locus *tg5e11a* (N12) to the model contributed significantly ($P = 0.0008$) to explaining yield in the inbreds (2.9% of the variance); the estimated average effect of substi-

tuting *RR* by *DD* at this locus represented −288.3 kg/ha. We note that a major flowering QTL mapped in the same genomic region and that we observed a negative correlation between yield and flowering time in the inbreds. Thus, this association is probably due to pleiotropy at the flowering time QTL rather than to linkage. We found additional evidence suggesting a pleiotropic effect of flowering time on yield of the inbred lines by completing a separate QTL analysis of yield for each year: in the Marnoo background in 1997, an average reduction in yield of 513 kg/ha ($P < 0.0001$) was observed in lines carrying *DD* instead of *RR* at *wg8g1b* (N2, near *VFN1*).

The QTL analysis for seed yield of the hybrids revealed that marker locus *wg2a11b* (N14) had a significant effect ($P = 0.0018$) in the combined analysis of both populations. It explained 2.6% of the variance and the average effect of substituting *TR* by *TD* at this locus represents 144.6 kg/ha (once the effects of backcross level, recurrent parent type, and translocation were included in the model). No other locus in the combined analysis had a significant effect. Analyzing each back-

TABLE 4

Marker loci significantly associated with seed weight and effect on seed weight of an allelic substitution at these loci once *wg3c5* (N15) has already been added to the model

Population	Marker locus	Linkage group[a]	Sign.	Variance explained (%)	Effect[b] (g/1000 seeds)
IBL *per se*					
Marnoo background	*wg2a6a*	N12	***	8.4	(0.815)[c]
	tg3c1a	N10	***	7.2	0.218
	wg4d10c	N17	****	8.8	0.256
Westar background	*wg3g9b*	Triplet 2	****	8.6	−0.288
	wgd10a	N3	***	6.7	0.224
	ec4h9	N5	****	13.8	0.365
IBL × Topas					
Marnoo background	*wg5b1c*	N12	***	5.4	−0.137
	ec5d5b	Unlinked	**	4.4	−0.110
	wg4d10c	N17	***	5.6	0.130
Westar background	*ec2d8a*	N3	****	10.5	0.144
	ec4c5d	Unlinked	**	5.5	−0.138
	wg1g5c	N1T	***	6.3	0.121

A separate analysis was done for each parental background. ** Significance level of 1%; ***significance level of 0.1%; ****significance level of 0.01%. Sign., significance level of marker locus.

[a] Nomenclature of Parkin *et al.* (1995)

[b] The effects indicated correspond to the substitution of *RR* by *DD* at the locus in the inbred lines and to the substitution of *TR* by *TD* in the hybrids.

[c] Only dominance had a significant effect at locus *wg2a6a* and 0.815 corresponds to *d*, half the estimated dominance value (with a standard error of 0.21).

ground separately, we found evidence for a second QTL for yield segregating in the Marnoo-derived hybrids near *wg1a4b* (N3). The estimated effect of substituting *TR* by *TD* at this locus represents 228.5 kg/ha ($P = 0.0009$) and it accounted for 6.6% of the variance. Substituting the *R* by the *D* alleles at both loci in the Marnoo-derived hybrids represents an estimated gain of 410.6 kg/ha.

Seed weight and oil content: Because there was a significant positive correlation between seed weight and seed yield, it is possible that Ceres could contribute a QTL allele that leads to smaller seed weight that would offset possible gains resulting from the introgression of winter germplasm. Indeed, marker locus *wg3c5* (N15) appears to be near such a QTL. The substitution of *TR* by *TD* at this locus is associated with a significant reduction in seed weight in the hybrids ($P < 0.0001$); on average it reduced seed weight by 0.140 g/1000 seeds and explained an additional 6.7% of the variance when added last to a model including effects due to parent, backcross, and translocation. It was also associated with a significant reduction of seed weight in the inbreds ($P = 0.0004$; −0.187 g/1000 seeds; 3.4% of the variance explained). The two backgrounds were analyzed separately to search for additional marker loci associated with seed weight, and three more QTL were detected in each hybrid background (Table 4). QTL analysis by background in the inbreds also yielded three additional QTL, some of which were in the same regions as the one mapped in the hybrids (N12 and N17 in the Marnoo

background, and N3 in the Westar background). At *wg2a6a*, we found evidence for a significant dominance effect, as the heterozygotes were associated with larger seed weight than either homozygote (Table 4). The four-locus models accounted for 31.8, 28.5, 28.2, and 26.6% of the variance in Marnoo inbreds and hybrids and Westar inbreds and hybrids, respectively.

The analysis of oil content revealed the location of only one putative QTL, near marker locus *tg1f8a* (on N1; $P = 0.0007$). The presence of the *DD* genotype at this locus was accompanied by an average increase in oil content of 1.1% (3.2% of the variance explained when added last to a model that already contained parent, backcrossing level, and main translocation).

CONCLUSIONS

Inbred backcross populations have been used to estimate the number of genes controlling a quantitative trait (Wehrhahn and Allard 1965) and to introgress unadapted germplasm into elite lines (Sullivan and Bliss 1983). When a molecular marker map was available from earlier generations such as a first backcross, populations similar to these were tested for significant marker-trait associations (Tanksley *et al.* 1996). In our study, it was shown that molecular marker data for inbred backcross populations can be used to perform linkage analyses and enrich a composite map. We also were able to test for marker-trait associations and gener-

ate multilocus models for several quantitative traits. The number of QTL found for each trait analyzed and the number of genes estimated by the method of Wehrhahn and Allard (1965) were of the same magnitude, and traits with a higher heritability allowed for the detection of a larger number of genes. This is probably due to the fact that QTL detection and the biometrical method of Wehrhahn and Allard (1965) are both constrained to the detection of genes with the largest effects [in contrast to the method of Castle and Wright (Castle 1921)]. The possibility of mapping QTL and simultaneously introgressing them into a desirable genetic background makes inbred backcross populations an attractive tool to study quantitative traits.

LITERATURE CITED

Alpert, K. B., and S. D. Tanksley, 1996 High-resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2.2*: a major fruit weight quantitative trait locus in tomato. Proc. Natl. Acad. Sci. USA **93:** 15503–15507.

Azanza, F., T. E. Young, D. Kim, S. D. Tanksley and J. A. Juvik, 1994 Characterization of the effect of introgressed segments of chromosome 7 and 10 from *Lycopersicon chmielewskii* on tomato soluble solids, pH, and yield. Theor. Appl. Genet. **87:** 965–972.

Baker, R. J., 1978 Evaluation of the inbred-backcross method for studying the genetics of continuous variation. Can J. Plant Sci. **58:** 7–12.

Banks, P. R., and W. D. Beversdorf, 1994 Self-incompatibility as a pollination control mechanism for spring oilseed rape, *Brassica napus* L. Euphytica **75:** 27–30.

Brandle, J. E., and P. B. E. McVetty, 1990 Geographical diversity, parental selection and heterosis in oilseed rape. Can. J. Plant Sci. **70:** 935–940.

Butruille, D. V., R. P. Guries and T. C. Osborn, 1999 Increasing yield of spring oilseed rape hybrids (*Brassica napus* L.) through introgression of winter germplasm. Crop Sci. (in press).

Castle, W. E., 1921 An improved method of estimating the number of genetic factors concerned in the case of blending inheritance. Science **54:** 223.

Collins, A., J. Teague, B. J. Keats and N. E. Morton, 1996 Linkage map integration. Genomics **36:** 157–162.

Démant, P., and A. A. M. Hart, 1986 Recombinant congenic strains—a new tool for analyzing genetic traits determined by more than one gene. Immunogenetics **24:** 416–422.

Diers, B. W., and T. C. Osborn, 1994 Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. Theor. Appl. Genet. **88:** 662–668.

Doebley, J., and A. Stec, 1991 Genetic analysis of the morphological differences between maize and teosinte. Genetics **129:** 285–295.

Dorweiller, J., A. Stec, J. Kermicle and J. Doebley, 1993 *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. Science **262:** 233–235.

Eshed, Y., and D. Zamir, 1994 A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. Euphytica **79:** 75–179.

Eshed, Y., and D. Zamir, 1996 Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics **143:** 1807–1817.

Ferreira, M. E., P. H. Williams and T. C. Osborn, 1994 RFLP mapping of *Brassica napus* using doubled haploid lines. Theor. Appl. Genet. **89:** 615–621.

Ferreira, M. E., J. Satagopan, B. S. Yandell, P. H. Williams and T. C. Osborn, 1995 Mapping loci controlling vernalization requirement and flowering time in *Brassica napus*. Theor. Appl. Genet. **90:** 727–732.

Graham, G. I., D. W. Wolff and C. W. Stuber, 1997 Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci. **37:** 1601–1610.

Howell, P. M., D. F. Marshall and D. J. Lydiate, 1996 Towards developing intervarietal substitution lines in *Brassica napus* using marker assisted selection. Genome **39:** 348–358.

Kosambi, D. D., 1944 The estimation of map distances from recombination values. Ann. Eugen. **12:** 172–175.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and artificial populations. Genomics **1:** 174–181.

Littell, R. C., G. A. Milliken, W. W. Stroup and R. D. Wolfinger, 1996 *SAS® System for Mixed Models.* SAS Institute, Cary, NC.

Liu, Y. G., N. Mitsukawa, C. Lister, C. Dean and R. F. Whittier, 1996 Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana*. Plant J. **10:** 733–736.

Mulitze, D. K., and R. J. Baker, 1985 Evaluation of biometrical methods for estimating the number of genes. 1. Effect of sample size. Theor. Appl. Genet. **69:** 553–558.

Osborn, T. C., D. C. Alexander and J. F. Fobes, 1987 Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. **73:** 350–356.

Osborn, T. C., C. Kole, I. A. P. Parkin, M. Kuiper, D. J. Lydiate *et al.*, 1997 Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. Genetics **146:** 1123–1129.

Parkin, I. A. P., A. G. Sharpe, D. J. Keith and D. J. Lydiate, 1995 Identification of the A and C genome of amphidiploid *Brassica napus* (oilseed rape). Genome **38:** 1122–1131.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln *et al.*, 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335:** 721–725.

Paterson, A. H., J. W. Deverna, B. Lanini and S. D. Tanksley, 1990 Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics **124:** 735–742.

Rau, G. A., C. E. McCulloch and M. A. Mutschler, 1994 Evaluation of parental-type classification for the inbred-backcross method of estimating gene number. J. Hered. **85:** 105–111.

Satagopan, J. M., B. S. Yandell, M. A. Newton and T. C. Osborn, 1996 A Baysian approach to detect quantitative trait loci using Markov Chain Monte Carlo. Genetics **144:** 805–816.

Sernyk, J. L., 1990 *Catalogue of Oilseed Rape Cultivars: 1990 Edition.* Agrigenetics Co., Madison, WI.

Sharpe, A. G., I. A. P. Parkin, D. J. Keith and D. J. Lydiate, 1995 Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome **38:** 1112–1121.

Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris and E. S. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics **132:** 823–839.

Sullivan, J. G., and F. A. Bliss, 1983 Expression of enhanced seed protein content in inbred backcross lines of common bean. J. Am. Soc. Hort. Sci. **108:** 787–791.

Tanksley, S. D., and J. C. Nelson, 1996 Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. **92:** 191–203.

Tanksley, S. D., S. Grandillo, T. M. Fulton, D. Zamir, Y. Eshed *et al.*, 1996 Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor. Appl. Genet. **92:** 213–224.

Teutonico, R. A., and T. C. Osborn, 1994 Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea* and *Arabidopsis thaliana*. Theor. Appl. Genet. **89:** 885–894.

Thormann, C. E., J. Romero, J. Mantet and T. C. Osborn, 1996 Mapping loci controlling the concentrations of erucic and linolenic acids in seed oil of *Brassica napus* L. Theor. Appl. Genet. **93:** 282–286.

Thurling, N., and L. D. Vijendra Das, 1979 Genetic control of

the pre-anthesis development of spring rape (*Brassica napus* L.). II. Identification of individual genes controlling developmental pattern. Aust. J. Agric. Res. **30:** 261–271.

Vos, P., R. Hogers, M. Bleake, M. Reijans, D. van de Lee *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:** 4407–4414.

Wehrhahn, C., and R. W. Allard, 1965 The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics **51:** 109–119.

APPENDIX A: ESTIMATION OF THE RECOMBINATION FREQUENCY BETWEEN TWO LOCI USING PAIRED INBRED BACKCROSS LINES

A1. Deduction of the formulas used to build the table of joint genotypic probabilities when the recurrent parent was homozygous at both marker loci: After calculating the probabilities of genotypes of the BC_1S_3 and of the BC_2S_2 separately conditioned upon the genotype of the BC_1 parents (Table A1), the final joint probability was calculated by adding over these conditional probabilities

$$P(BC_1S_3, BC_2S_2) = (1 - \theta)/2$$
$$\times [P(BC_1S_3 \mid BC_1 = AABB)$$
$$\times P(BC_2S_2 \mid BC_1 = AABB)$$
$$+ P(BC_1S_3 \mid BC_1 = AaBb)$$
$$\times P(BC_2S_2 \mid BC_1 = AaBb)] + \theta/2$$
$$\times [P(BC_1S_3 \mid BC_1 = AaBB)$$
$$\times P(BC_2S_2 \mid BC_1 = AaBB)$$
$$+ P(BC_1S_3 \mid BC_1 = AABb)$$
$$\times P(BC_2S_2 \mid BC_1 = AABb)].$$

The likelihood of the recombination fraction between two loci being θ was obtained by multiplying the joint probabilities (of observing the data given θ) of all informative BC_1S_3-BC_2S_2 pairs. The value of θ that resulted in the highest likelihood was chosen as the estimate of recombination fraction between the two loci.

A2. Pedigree configuration encountered as a function of the genotype of the recurrent parents at a marker locus: Not all loci were homozygous in the recurrent parents. For this reason, and given the pedigree structure used, there are 11 possible configurations of the parental genotypes at a locus (Table A2). Thus, exact determination of all possible two-point comparisons in these populations requires $11 \times 11 = 121$ contingency tables of genotypic probabilities of the pairs of IBLs. Knowledge of the phase in the recurrent parent (when both loci are heterozygous in the dominant parent) is important for 110 of these, hence the total number of tables is actually 231 (121 + 110). As we had some pairs of IBLs for which one member of the pair was not correctly selfed, these 231 tables were redone three

TABLE A1

Expected probabilities of genotypes in IBLs conditioned upon the genotype of the BC_1 parent

Genotype of BC_1 (probability)	Possible genotypes of the IBLs	Conditional probability of a BC_1S_3 IBL [i.e., $P(BC_1S_3 \mid BC_1)$]	Conditional probability of a BC_2S_2 IBL [i.e., $P(BC_2S_2 \mid BC_1)$]
$AABB\ [(1 - \theta)/2]$[a]	AABB	1	1
$AaBB\ (\theta/2)$	AABB	7/16	11/16
	AaBB	1/8	1/8
	aaBB	7/16	3/16
$AABb\ (\theta/2)$	AABB	7/16	11/16
	AABb	1/8	1/8
	AAbb	7/16	3/16
$AaBb\ [(1 - \theta)/2]$	AABB	$(7 - 14\theta + 19\theta^2 - 20\theta^3 + 18\theta^4 - 12\theta^5 + 4\theta^6)/16$	$(11 - 11\theta + 12\theta^2 - 10\theta^3 + 6\theta^4 - 2\theta^5)/16$
	AABb	$(1 - \theta)\theta(3 - 6\theta + 10\theta^2 - 8\theta^3 + 4\theta^4)/8$	$\theta(3 - 6\theta + 8\theta^2 - 6\theta^3 + 2\theta^4)/8$
	AAbb	$\theta(8 - \theta - 12\theta^2 + 18\theta^3 - 12\theta^4 + 4\theta^5)/16$	$\theta(5 - 6\theta^2 + 6\theta^3 - 2\theta^4)/16$
	AaBB	$(1 - \theta)\theta(3 - 6\theta + 10\theta^2 - 8\theta^3 + 4\theta^4)/8$	$\theta(3 - 6\theta + 8\theta^2 - 6\theta^3 + 2\theta^4)/8$
	AaBb	$(1 - 2\theta + 2\theta^2)^3/8$	$(1 - \theta)(1 - 2\theta + 2\theta^2)^2/8$
	Aabb	$(1 - \theta)\theta(3 - 6\theta + 10\theta^2 - 8\theta^3 + 4\theta^4)/8$	$(1 - \theta)^2\theta(1 - \theta + \theta^2)/4$
	aaBB	$\theta(8 - \theta - 12\theta^2 + 18\theta^3 - 12\theta^4 + 4\theta^5)/16$	$\theta(5 - 6\theta^2 + 6\theta^3 - 2\theta^4)/16$
	aaBb	$(1 - \theta)\theta(3 - 6\theta + 10\theta^2 - 8\theta^3 + 4\theta^4)/8$	$(1 - \theta)^2\theta(1 - \theta + \theta^2)/4$
	aabb	$(7 - 14\theta + 19\theta^2 - 20\theta^3 + 18\theta^4 - 12\theta^5 + 4\theta^6)/16$	$(1 - \theta)(3 - 6\theta + 6\theta^2 - 4\theta^3 + 2\theta^4)/16$

[a] θ, recombination fraction.

Possible marker locus configurations in parents of IBL populations and number of loci scored belonging to each class

Individual	Possible genotypic combinations										
	1	2	3	4	5	6	7	8	9	10	11
Donor parent	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD	DD
Recurrent parent	RR	RD	RD	RD	RD	RD	RD	RD	RD	RD	RD
Recurrent parent S_1	RR	RR	RD	RD	RD	DD	RR	RD	RD	RD	DD
Recurrent parent S_2	RR	RR	RR	RD	DD	DD	RR	RR	RD	DD	DD
F_1	RD	RD	RD	RD	RD	RD	DD	DD	DD	DD	DD
Population	No. of marker loci for each combination										
	1	2	3	4	5	6	7	8	9	10	11
Marnoo subpop1	104	4	1	2	0	2	2	3	0	3	1
Marnoo subpop2	104	1	1	4	1	2	1	2	4	1	1
Westar subpop1	121	0	0	0	1	0	0	0	0	0	0
Westar subpop2	121	0	0	0	0	0	0	0	1	0	0

times ($231 \times 3 = 693$) for each of the possible configuration of pairs [(BC_1S_3, BC_2S_2), (BC_1S_2, BC_2S_2), and (BC_1S_3, BC_2S_1)]. These tables are nine rows by nine columns matrices, based on the number of genotypes possible at two loci in one line of the IBL pair.

APPENDIX B: SOFTWARE USED TO CALCULATE PAIRWISE INFORMATION FROM THE RAW MOLECULAR MARKER DATA

MAP+ (Collins *et al.* 1996) requires pairwise information to produce a genetic map of each linkage group. We wrote two programs in ANSI/ISO C that could use the raw marker data and extract the complete pairwise information. These programs use probability tables in an iterative fashion to conduct a grid search at a final precision of 0.1%.

The first program reads a file containing the raw data, information on phase of loci heterozygous in the recurrent parent (mostly in the Marnoo population), and information about the level to which a line was selfed (BC_1S_3 lines were sometimes only BC_1S_2 and BC_2S_2 lines were sometimes only BC_2S_1). Data from the four IBL populations are combined. In the first iteration, the table is built assuming that all loci are homozygous in the parents and that all the lines are at the ideal selfing level. The second iteration corrects the result obtained where those assumptions are violated and scans (for each pair of loci), from 0–50% recombination at 1% intervals, for the recombination fraction that yields the highest LOD score. These final results are then pooled with the pairwise information from the population studied by Ferreira *et al.* (1994); this information is easily obtained through the direct counting of recombinants, as doubled haploids yield the same information as F_1 gametes. The resulting data (*i.e.*, the

pairwise recombination frequency and LOD scores pooled over the three populations) are used to split the loci into subsets that are the putative linkage groups. The information on linkage group is used by the second program and to write a preliminary "job" file required by MAP+.

The second program is essentially identical to the first one but it uses the information on linkage groups to recalculate the pairwise information between loci within the same linkage group. This is done separately for each pair of IBL populations (WBC_1S_3, WBC_2S_2) and (MBC_1S_3, MBC_2S_2), and for the doubled-haploid population. This will allow MAP+ to test for homogeneity of recombination fraction across populations. The scanning is done at 0.1% intervals. The final outputs are three tables with pairwise information combined and formatted to be used as an input file by MAP+.

The correctness of the program was tested by confirming that the sum of the probabilities in the contingency tables added to 1.0000 for any percentage of recombination and by comparing the probabilities calculated by these programs with the probabilities expected when the recombination is 0 and 50%. The good agreement between the map obtained with MAP+ and the one from Ferreira *et al.* (1994) also helps validate those programs.

APPENDIX C: SOFTWARE TO CALCULATE GENOTYPIC PROBABILITIES AT LOCI

A third program was written (in ANSI/ISO C) to permit the use of marker locus information in the QTL analysis. It uses the information available at a marker locus and at surrounding loci to assign genotypic probabilities at this locus for each IBL. If the information at

this locus is complete (in the case of a fully informative codominant marker), the genotypic probability is one for the genotype scored (*e.g.*, *DD*) and zero for the remaining two (*e.g.*, *DR* and *RR*). Otherwise, in the case of a dominant locus or missing data, these genotypic probabilities are estimated on the basis of the partial information at this locus in the IBL under scrutiny (*e.g.*, $R\_$ in the BC_1S_3), on information at this locus on the paired IBL (*e.g.*, *DD* in the BC_2S_2), and on information at flanking marker loci. On the basis of the hypothesis of no interference, if the closest marker locus on one side was completely informative, there is no gain in knowing genotypes of loci farther away, and this property was used to reduce computing time. Absence of interference was also assumed to simplify multilocus probabilities into products of two-locus probabilities. That is, if G_A, G_B, G_C, G_D, and G_E represent the complete genotypes (both alleles and the phase are known) at the five consecutive loci A, B, C, D, and E, under the hypothesis of no interference we have

$$P(G_A, G_B, G_C, G_D, G_E) = [P(G_A, G_B) P(G_B, G_C) P(G_C, G_D)$$
$$P(G_D, G_E)] / [P(G_B) P(G_C) P(G_D)].$$

This equation is at the base of a nested loop structure used in this program.

   This program was tested by checking if the probabilities added to one by comparing probabilities calculated for loci for which no marker information was available within the 43-cM window to the probabilities calculated on the basis of information on the pedigree only and by checking for inconsistencies in the resulting matrix of probabilities.