Comparative Quantitative Trait Loci Mapping of Aliphatic, Indolic and Benzylic Glucosinolate Production in *Arabidopsis thaliana* Leaves and Seeds

Daniel J. Kliebenstein,[1] Jonathan Gershenzon and Thomas Mitchell-Olds

Max Planck Institute for Chemical Ecology, 07745 Jena, Germany

Manuscript received February 21, 2001
Accepted for publication July 3, 2001

ABSTRACT

Secondary metabolites are a diverse set of plant compounds believed to have numerous functions in plant-environment interactions. Despite this importance, little is known about the regulation of secondary metabolite accumulation. We are studying the regulation of glucosinolates, a large group of secondary metabolites, in Arabidopsis to investigate how secondary metabolism is controlled. We utilized L*er* and Cvi, two ecotypes of Arabidopsis that have striking differences in both the types and amounts of glucosinolates that accumulate in the seeds and leaves. QTL analysis identified six loci determining total aliphatic glucosinolate accumulation, six loci controlling total indolic glucosinolate concentration, and three loci regulating benzylic glucosinolate levels. Our results show that two of the loci controlling total aliphatic glucosinolates map to biosynthetic loci that interact epistatically to regulate aliphatic glucosinolate accumulation. In addition to the six loci regulating total indolic glucosinolate concentration, mapping of QTL for the individual indolic glucosinolates identified five additional loci that were specific to subsets of the indolic glucosinolates. These data show that there are a large number of variable loci controlling glucosinolate accumulation in *Arabidopsis thaliana*.

THE sessile nature of plants forces them to cope directly with environmental changes rather than escape to more favorable sites. Secondary metabolites are a large and diverse set of molecules that allow plants to respond to these environmental challenges, which include insect herbivory, pathogen attack, UV radiation, and drought. There are currently >100,000 known secondary metabolites, which probably represent <10% of the total in nature (Wink 1988). This large chemical diversity presumably arises from a vast set of biosynthetic enzymes. Although numerous studies have investigated the enzymes and regulatory factors controlling biosynthesis of secondary metabolites, little is known about the genetics controlling quantitative and qualitative natural variation in secondary chemistry. To address this question, we are studying glucosinolate production in *Arabidopsis thaliana*.

Glucosinolates are amino acid-derived thioglycosides. Glucosinolates and their hydrolyzing agent, myrosinase, are spatially separated within plant cells (Bones and Rossiter 1996). When the cell is disrupted, myrosinase cleaves the sugar from the glucosinolate, and a series of toxic compounds are released. These toxins include nitriles, isothiocyanates, oxozaladines, and epithioalkanes. Glucosinolates and their breakdown products have been shown to alter insect herbivory (Siemens and Mitchell-Olds 1996; Stowe 1998; Li *et al.* 2000), nematode survival (Donkin *et al.* 1995), and fungal resistance (Mithen *et al.* 1986; Manici *et al.* 1997). Additionally, certain glucosinolates have anticarcinogenic activities in mammals (Verhoeven *et al.* 1997).

Arabidopsis utilizes a three-part biosynthetic pathway to produce glucosinolates from methionine, tryptophan, and phenylalanine (Hogge *et al.* 1988; Halkier and Du 1997). The first step in this pathway is the elongation of a protein amino acid (Figure 1). This is accomplished by the deamination of the amino acid to the 2-oxo acid and subsequent methylene group addition (de Quiros *et al.* 2000; Graser *et al.* 2000; Figure 1). The 2-oxo acid can then be reaminated to the elongated amino acid or undergo further elongation (Figure 1). The amino acid enters the general pathway of glucosinolate biosynthesis by conversion to an aldoxime via one of a series of cytochrome P450 oxygenases (Kahn *et al.* 1999; Wittstock and Halkier 2000). The resulting amino acid is then converted to a glucosinolate by three additional steps. Finally, a number of amino acid-specific side chain modifications can occur (Kliebenstein *et al.* 2001b; Figure 1). Several of the genes controlling these steps have been mapped and/or cloned in *A. thaliana*, generating a detailed model of glucosinolate biosynthesis. This provides a solid base from which to interpret natural variation in glucosinolates.

More than 35 different glucosinolates have been documented in the species *A. thaliana*, with 25 of these coming from differential elongation and side-chain modifi-

Corresponding author: Thomas Mitchell-Olds, Max Planck Institute for Chemical Ecology, Carl Zeiss Promenade 10, 07745 Jena, Germany. E-mail: tmo@ice.mpg.de

[1] *Present address:* Department of Vegetable Crops, 148 Asmundson Hall, University of California, Davis, CA 95616.

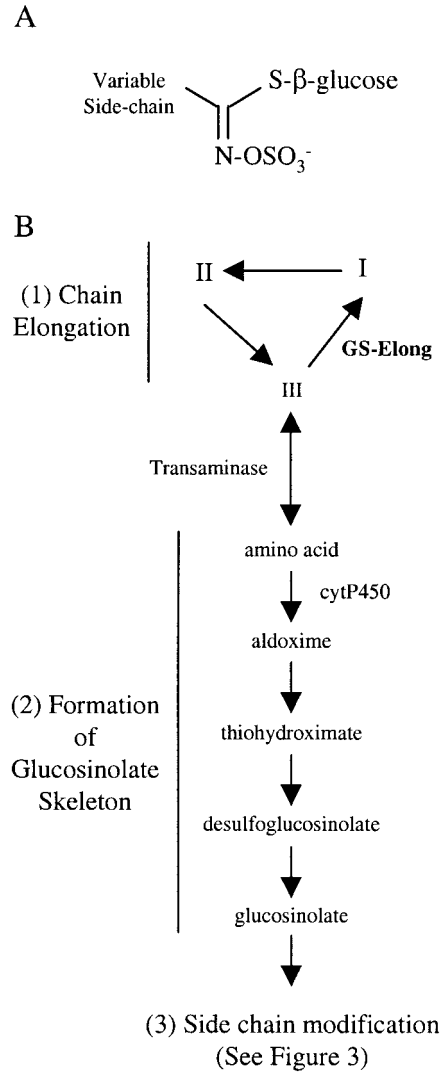

FIGURE 1.—Glucosinolate biosynthetic pathway. I, 2-alkylmalic acid; II, 3-alkylmalic acid; III, 2-oxo acid. (A) Basic glucosinolate structure. (B) Outline of the pathway, which can be divided into three parts: (1) elongation of the amino acid side chain; (2) formation of the basic glucosinolate skeleton; and (3) further side chain modification. Each chain elongation cycle adds an additional methylene group (GRASER *et al.* 2000).

cation of methionine (MITHEN and CAMPOS 1996; KLIEBENSTEIN *et al.* 2001b). In contrast to the large number of different glucosinolates present in the species as a whole, the leaves of individual Arabidopsis plants contain only 7–14 different glucosinolates, with 4–11 of these coming from methionine. This qualitative variation in methionine-derived glucosinolates can be explained by natural variation at four biosynthetic loci (KLIEBENSTEIN *et al.* 2001b). In contrast to qualitative variation in Arabidopsis, little is known about the factors controlling quantitative variation in aliphatic, indolic, or benzylic glucosinolates (KLIEBENSTEIN *et al.* 2001b).

To identify loci controlling quantitative variation in Arabidopsis glucosinolates we initiated a quantitative trait locus (QTL) mapping experiment using the Landsberg *erecta* (L*er*) × Cape Verdi Islands (Cvi) recombinant inbred lines (RILs; ALONSO-BLANCO *et al.* 1998). Previous studies examining glucosinolates QTL focused on the accumulation of seed glucosinolates in *Brassica napus*. These studies identified four to six QTL that controlled aliphatic glucosinolate concentration (TOROSER *et al.* 1995; UZUNOVA *et al.* 1995). However, little is known about the genes underlying these QTL or about QTL that control nonaliphatic glucosinolates or leaf glucosinolates. In this article, we report on QTL mapping experiments studying aliphatic, indolic, and benzylic glucosinolate concentration in both the leaves and seeds of the *A. thaliana* L*er* × Cvi RILs (ALONSO-BLANCO *et al.* 1998). These QTL are discussed in relation to previously described Arabidopsis glucosinolate biosynthetic loci (DE QUIROS *et al.* 2000; KLIEBENSTEIN *et al.* 2001a). We also investigate the coregulation of different glucosinolate classes in the two different tissues. Finally, we describe the identification of QTL that control the accumulation of specific glucosinolates.

MATERIALS AND METHODS

Plant growth: All seed stocks were obtained from the Arabidopsis Biological Resource Center (http://aims.cps.msu.edu/aims/). Five plants per L*er* × Cvi recombinant inbred lines were planted in a 6-cm pot containing a standard vermiculite/soil mix. The pots were placed in flats of 48 in a randomized design with one pot of L*er* and one pot of Cvi per flat. This was repeated three times with independently randomized designs. The plants were allowed to grow at 26° for 3 weeks postgermination. They were placed 10 in. from four 60-W Cool White Delux GE lightbulbs in a 12-hr light/12-hr dark photoperiod. At this time, 10 leaves were harvested for glucosinolate extraction. The same plants were allowed to continue to complete seed maturation. The seeds were then combined from all the plants in each pot and kept separate from the seeds harvested from the other plantings. Seed glucosinolates were analyzed independently from each of the three replicates per recombinant inbred line and from the 10 replicates of L*er* and Cvi parental lines.

Sample preparation and HPLC: Samples were extracted and analyzed by HPLC as previously described (KLIEBENSTEIN *et al.* 2001a). Specific glucosinolates were identified by comparison of retention times and UV absorption spectra with purified standards. All glucosinolate absorption data (measured at 229 nm) were converted to micromoles per gram dry weight (DWT) using response factors determined from the purified standards for each of the glucosinolates (KLIEBENSTEIN *et al.* 2001a).

Glucosinolate traits: Glucosinolate traits were generated using data on individual compounds. TALIPH is the sum of all methionine-derived (aliphatic) glucosinolates. TIND is the sum of all tryptophan-derived or indolic glucosinolates. Benzyl is the sum of all phenylalanine-derived or benzyl glucosinolates. MO to I3M = (1-methoxyI3M + 4-methoxyI3M)/I3M.

QTL mapping and statistics: The least-squares mean for each recombinant inbred line for the L*er* × Cvi leaf and seed samples was obtained from Systat, utilizing each of the three replicates in an analysis of variance using flat as a categorical factor to control for possible heterogeneous environmental conditions. Relative glucosinolate concentrations for all 162

lines were then used for QTL mapping, utilizing QTLcartographer version 1.13 with both linear interval mapping and composite interval mapping (Basten *et al.* 1999). We determined significance thresholds for each trait by doing 500 permutations to estimate the 0.05 significance level. Marker data employed 99 amplified fragment length polymorphism (AFLP) markers from Alonso-Blanco *et al.* (1999) and 5 microsatellite markers that were previously described (Kliebenstein *et al.* 2001a). Mapping of the markers for both RI lines employed Mapmaker version 3 (Lander *et al.* 1987).

For Table 2, heritability was estimated by ANOVA analysis. The R^2 was determined from two different ANOVA models:

1. Glucosinolate = constant + line + flat
2. Glucosinolate = constant + flat

The heritability of the trait due to line effects was then estimated as the difference in R^2 values obtained between the two models (heritability = $R_1^2 - R_2^2$).

For Tables 3 and 4, epistatic interactions among QTL were analyzed by ANOVA using the genetic markers showing the largest statistical control for each QTL. The models used are listed below:

Table 3: Mean leaf aliphatic glucosinolates = constant + *GS-AOP* + *EC198L* + *GS-ELONG* + *GS-AOP* × *GS-ELONG* + *EC198L* × *GS-ELONG* + *GS-AOP* × *EC198L* + *GS-AOP* × *EC198L* × *GS-ELONG*

Table 4: Mean seed benzylic glucosinolate = constant + *GS-ELONG* + *AD182C* + *GB120C* + *AD182C* × *GS-ELONG* + *AD182C* × *GB120C* + *GS-ELONG* × *GB120C*

TABLE 1

Identified glucosinolates

	L*er*		Cvi		RILs	
Compound	L[a]	S[b]	L	S	L	S
3-Methylthiopropyl		+			+	+
3-Methylsulfinylpropyl				+	+	+
3-Hydroxypropyl	+	+		+	+	+
3-Benzoyloxypropyl		+		+		+
Allyl			+	+	+	+
4-Methylthiobutyl		+		+	+	+
4-Methylsulfinylbutyl				+	+	+
4-Hydroxybutyl		+		+		+
4-Benzoyloxybutyl		+		+		+
3-Butenyl			+	+	+	+
2-Hydroxy-3-butenyl					+	+
2-Benzoyloxy-3-butenyl						+
4-Pentenyl			+	+	+	+
7-Methylthioheptyl	+	+	+	+	+	+
7-Methylsulfinylheptyl		+	+	+	+	+
8-Methylthiooctyl	+	+	+	+	+	+
8-Methylsulfinyloctyl	+	+	+	+	+	+
Indolyl-3-methyl (I3M)	+	+	+	+	+	+
4-HydroxyI3M	+		+		+	
1-MethoxyI3M	+	+	+	+	+	+
4-MethoxyI3M	+		+		+	
Phenylethyl		+		+		+

[a] + indicates that glucosinolate is present in the leaf.
[b] + indicates that glucosinolate is present in the seed.

RESULTS

Qualitative aliphatic glucosinolate variation in the L*er* × Cvi RILs: The L*er* and Cvi *A. thaliana* ecotypes accumulate different glucosinolate types. L*er* leaves contain predominantly 3-hydroxypropyl and 8-methylsulfinyloctyl glucosinolate while Cvi leaves have principally allyl and 3-butenyl glucosinolate (Table 1 and Figure 2). This difference in glucosinolate profile is controlled by the previously described *GS-AOP* and *GS-Elong* loci (Kliebenstein *et al.* 2001a,b). *GS-AOP* controls the difference between hydroxyalkyl (*i.e.*, hydroxypropyl) and alkenyl (*i.e.*, allyl and 3-butenyl) glucosinolates while the *GS-Elong* locus controls C_3 (*i.e.*, allyl and hydroxypropyl) *vs.* C_4 (*i.e.*, 3-butenyl) glucosinolate production (Figure 3). These loci are semidominant because the glucosinolate profile of the F_1 progeny from a L*er* × Cvi cross contains all glucosinolates present in both parents (Figure 2). About one-half of the L*er* × Cvi RILs display glucosinolate profiles similar to either the L*er* or Cvi parents (Figure 3). However, genetic segregation has generated three new glucosinolate profiles that are not found in either parent (Figure 3). For example, 2-hydroxy-3-butenyl glucosinolate is not found in either parent (Table 1 and Figure 3). This new glucosinolate is produced by the introduction of a hydroxyl group at the C-2 of 3-butenyl glucosinolate (Rossiter *et al.* 1990). As Cvi is incapable of this reaction, L*er* appears to contain an enzyme able to catalyze this reaction, although it does not contain the 3-butenyl precursor (Table 1 and Figure 2). In comparison to leaves, Cvi seeds accumulate low levels of 3-hydroxypropyl and 4-hydroxybutyl glucosinolate in addition to allyl and 3-butenyl glucosinolate (Table 1). Further, seeds of both ecotypes contain high levels of benzoic acid ester derivatives of all hydroxyalkyl and hydroxyalkenyl glucosinolates (Table 1; Kliebenstein *et al.* 2001b).

Quantitative glucosinolate variation between L*er* and Cvi: In addition to the qualitative variation discussed above, L*er* and Cvi exhibit quantitative glucosinolate variation (Figure 2 and Table 2). Cvi leaves accumulate nearly 8-fold more total aliphatic glucosinolates than *Ler* leaves (Figure 2 and Table 2). Leaves of L*er* × Cvi F_1 progeny contain intermediate glucosinolate levels (Figure 2). In addition to the large difference between the parents, the RILs displayed a 39-fold difference between the high and low lines for leaf aliphatic glucosinolates (Table 2). Thus, both ecotypes appear to have increasing and decreasing alleles, which allow for transgressive segregation in total quantities of glucosinolates.

Whereas L*er* and Cvi had significantly different leaf aliphatic glucosinolate levels, they have nearly identical seed aliphatic glucosinolate accumulation (Figure 2 and Table 2). In contrast to the similarity between the two parents, the high and low L*er* × Cvi RILs had a nearly threefold difference in total seed aliphatic glucosinolate concentration (Table 2). This indicates that the similar

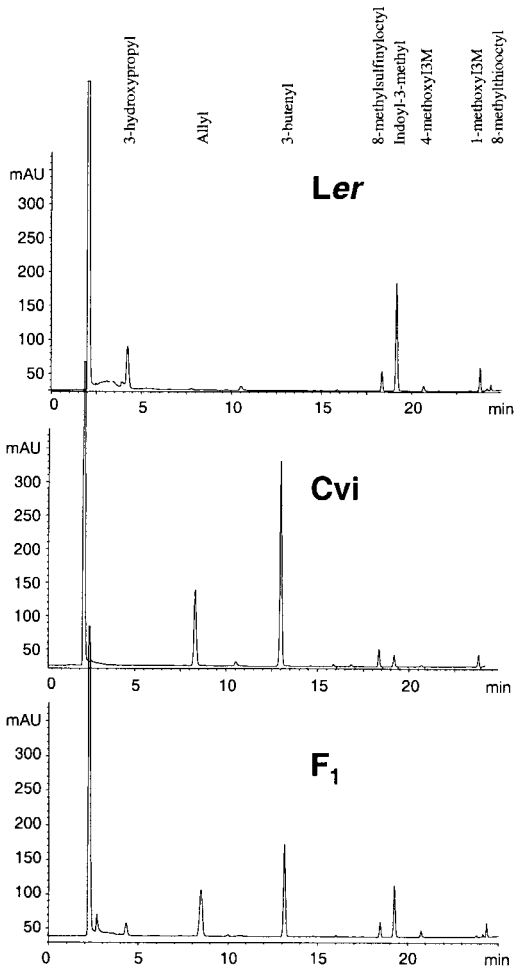

FIGURE 2.—HPLC chromatograms of compounds present in the leaves of L*er*, Cvi, and F_1 progeny. Shown are HPLC chromatograms monitored at 229 nm of samples from L*er*, Cvi, and L*er* × Cvi F_1 progeny prepared from equal amounts of 2-week-old rosette tissue. The major glucosinolate peaks are labeled at the top of the chromatogram.

levels of aliphatic glucosinolates in seeds hide genetic differences between L*er* and Cvi for the control of total aliphatic seed glucosinolates.

In contrast to methionine-derived aliphatic glucosinolates, where Cvi is higher than L*er*, L*er* has 2- to 3-fold higher levels of both indolic and benzylic glucosinolates in all tissues (Figure 2 and Table 2). The F_1 progeny showed that leaf indolic glucosinolate accumulation was not under the control of a single dominant or recessive locus (Figure 2). Further, the RILs display a striking amount of transgressive segregation for all three traits. This ranges from a 34-fold difference between high and low lines for seed benzylic glucosinolate accumulation to a 13-fold difference between high and low RILs for leaf indolic glucosinolate concentration (Table 2). Thus, there is substantial genetic variation for aromatic glucosinolate accumulation.

The heritability for most traits is high, with leaf total aliphatic glucosinolates being the highest at 0.81 (Table 2). This suggests that genetic factors segregating between the lines control ~80% of the variation seen in the experiment. In contrast, seed total aliphatic glucosinolates had a heritability of only 0.40 (Table 2). This indicates that genetics contributes ~40% of the variation while environmental and experimental factors contribute 60% of the variation. Even compounds in low abundance like 8-methylthiooctyl in the leaves (see Figure 2) had a heritability of 0.63. Thereby, the detection of even low-concentration compounds is highly repeatable and their accumulation is under genetic control.

Coregulation of glucosinolate accumulation between leaves and seeds: Previous work showed a significant correlation between the levels of leaf and seed aliphatic glucosinolates in 35 different Arabidopsis ecotypes (KLIEBENSTEIN *et al.* 2001b). We further tested this by examining the correlation of leaf and seed aliphatic glucosinolate concentration in the leaves and seeds of the 162 L*er* × Cvi RILs. This analysis showed a positive correlation between aliphatic glucosinolates in the two different tissues (Figure 4, $r = 0.19$, $P = 0.015$, $n = 162$). Further, concentration of aliphatic glucosinolates was always higher in seeds than in leaves. This suggests that the glucosinolate biosynthesis in the leaf may determine the lower limit for seed aliphatic glucosinolate accumulation. Support for this theory comes from the observation that 35 Arabidopsis ecotypes displayed a similar relationship between leaf and seed glucosinolates (KLIEBENSTEIN *et al.* 2001b). Additionally, there was a positive correlation between indolic glucosinolate accumulation in the two different tissues (Figure 4, $r = 0.40$, $P < 0.001$, $n = 162$).

Mapping QTL controlling total aliphatic glucosinolate accumulation: To investigate the genetic control of quantitative variation in Arabidopsis glucosinolates, we utilized the mean total leaf aliphatic glucosinolate concentration for each of the 162 RILs to map QTL. This analysis identified three QTL that controlled 68% of the variation in leaf aliphatic glucosinolate levels (Figure 5). Two of these QTL are linked with the previously known glucosinolate biosynthetic loci, *GS-Elong* and *GS-AOP*. This was expected, as both loci were previously shown to control the concentration of aliphatic glucosinolates (MITHEN *et al.* 1995; KLIEBENSTEIN *et al.* 2001a). Fine-scale mapping showed that the *GS-Elong* and *GS-AOP* QTL have LOD maxima at markers that are physically located within the *GS-Elong* and *GS-AOP* genes (D. J. KLIEBENSTEIN and T. MITCHELL-OLDS, unpublished data; DE QUIROS *et al.* 2000; KLIEBENSTEIN *et al.* 2001a). The second strongest QTL controlling leaf aliphatic glucosinolates is linked with the AFLP marker *EC198L* and is a previously unidentified Arabidopsis glucosinolate locus. In contrast to *GS-Elong* and *GS-AOP*, which control both the type and amount of aliphatic glucosinolate produced, there is no correlation between *EC198L* and the presence or absence of any specific glucosinolate (D. J. KLIEBENSTEIN and T. MITCHELL-OLDS, unpub-

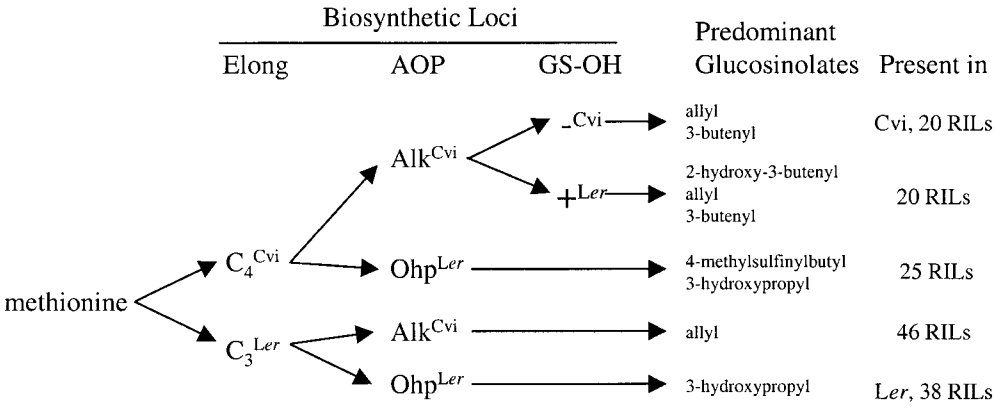

FIGURE 3.—Genetic control of glucosinolate profiles in L*er* × Cvi RILs. Shown is a flow chart of how the three variable glucosinolate biosynthetic loci (*GS-AOP*, *GS-Elong*, and GS-OH) alter accumulation of the predominant aliphatic glucosinolates in leaves of L*er* × Cvi RILs. Also listed is the number of RILs displaying each pattern of glucosinolate accumulation. Only the 149 lines with sufficent aliphatic glucosinolate levels to accurately score the genotype are shown. Seed glucosinolates are similar except that hydroxy glucosinolates are converted to benzoyloxy glucosinolates and 4-methylsulfinylbutyl glucosinolate is converted to 4-hydroxybutyl and 4-benzoyloxybutyl glucosinolate.

lished data). This suggests that the *EC198L* QTL is a single locus or several closely linked loci that controls the production of leaf aliphatic glucosinolates irrespective of type.

To compare the factors regulating leaf and seed aliphatic glucosinolates, we mapped QTL controlling the accumulation of seed aliphatic glucosinolates. This analysis also uncovered three QTL, which were all seed specific (Figure 5). Together, these QTL explain 38% of the variation in seed aliphatic glucosinolate concentration. An analysis of the presence of specific glucosinolates showed that the QTL near *Erecta* on chromosome II is tightly linked with the presence of 2-hydroxy-3-butenyl glucosinolate (2.5 cM between *Erecta* and the 2-hydroxy-3-butenyl locus). This suggests that this locus may regulate the formation of 2-hydroxy-3-butenyl glucosinolate in addition to regulating total seed aliphatic glucosinolate levels. In contrast, the QTL associated with GD86L on chromosome I and BF168L on chromosome V did not correlate with the presence or absence of any specific glucosinolates.

Epistasis of QTL controlling aliphatic glucosinolate concentration: The observation that two of the three QTL controlling leaf aliphatic glucosinolate levels were linked with glucosinolate biosynthetic loci suggested that these loci might interact to control leaf aliphatic glucosinolate accumulation. We utilized ANOVA to test this possibility. This showed that *GS-Elong* and *GS-AOP* display a large and highly significant interaction controlling leaf aliphatic glucosinolate accumulation (Table 3 and Figure 6). This interaction is such that the presence of a L*er* homozygote at *GS-Elong* or *GS-AOP* functions to decrease leaf aliphatic glucosinolate levels (Figure 6). In addition to interacting with *GS-AOP*, *GS-Elong* also interacts with the *EC198L* QTL (Table 3 and Figure 6). The same pattern of interaction is present with the RILs containing the Cvi alleles at *GS-Elong* and *EC198L*, having about twofold higher glucosinolate concentrations than the RILs with the other three allele combinations (Figure 6). In addition to the two-way interactions, *GS-Elong*, *GS-AOP*, and *EC198L* display a significant three-way interaction (Table 3). These interactions suggest

TABLE 2

Means for total glucosinolate traits

			RI Lines			
Trait	L*er*	Cvi	Mean	Min	Max	Heritability
TALIPHL[a]	1.7	12.3	25.6	0.5	19.9	0.81
TALIPHS[b]	22.7	22.7	25.6	12.6	34.6	0.40
TINDL[c]	1.05	0.43	0.77	0.19	2.51	0.67
TINDS[d]	0.17	0.13	0.14	0.02	0.52	0.59
BenzylS[e]	0.68	0.24	0.42	0.03	1.02	0.72

Values are given in μmol $gDWT^{-1}$. Ten replicates of L*er* and Cvi and three replicates of each RIL were analyzed by HPLC for glucosinolate concentration. The mean given for the RILs is the mean of all 162 lines analyzed in triplicate.

[a] TALIPHL is the sum of all aliphatic glucosinolates in the leaf.
[b] TALIPHS is the sum of all aliphatic glucosinolates in the seed.
[c] TINDL is the sum of all indolic glucosinolates in the leaf.
[d] TINDS is the sum of all indolic glucosinolates in the seed.
[e] BenzylS is the sum of all phenylalanine-derived glucosinolates in the seed.

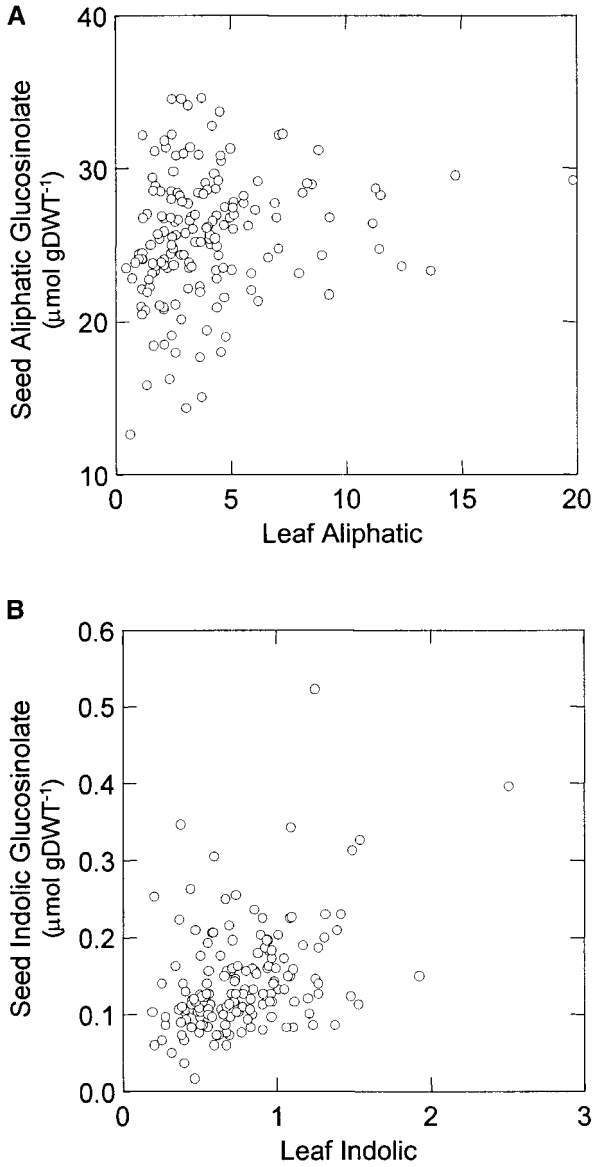

FIGURE 4.—Correlation between leaf and seed glucosinolate concentration. Scatter plots are shown comparing the accumulation of leaf and seed glucosinolates. (A) The relationship between aliphatic glucosinolate in the leaf (LTALIPH) and seed (STALIPH). (B) The relationship between indolic glucosinolate in the leaf (LTIND) and seed (STIND).

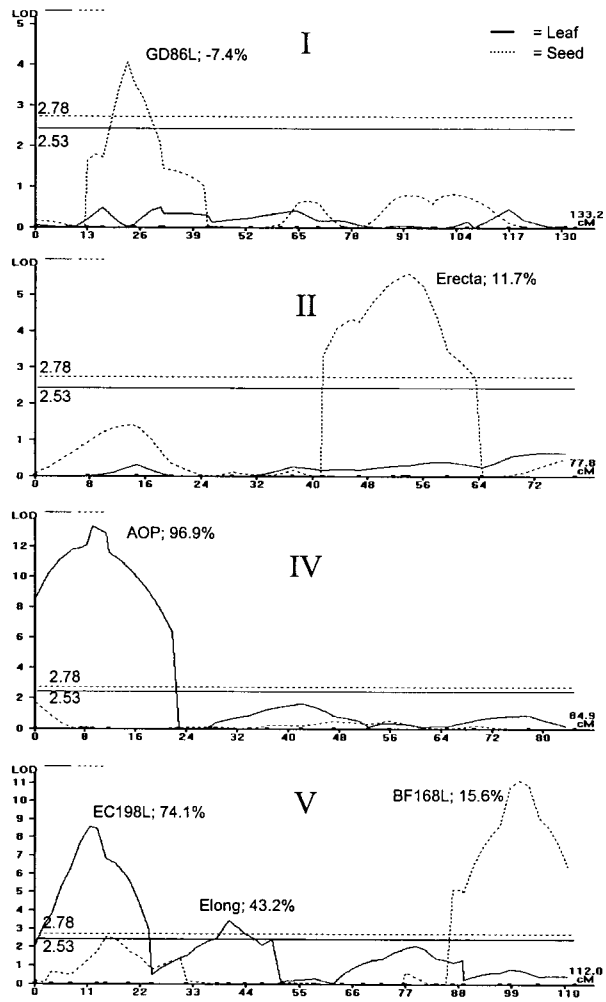

FIGURE 5.—QTL controlling aliphatic glucosinolate concentration in the leaves and seeds of the L*er* × Cvi RILs. LOD plots of QTL for the accumulation of total aliphatic glucosinolates are shown. Total aliphatic glucosinolate accumulation was determined by summing all of the individual methionine-derived glucosinolates. Numbers are the percentages of difference in mean glucosinolate concentration between lines homozygous for the Cvi allele and lines homozygous for the L*er* allele at the listed markers. The mean from the homozygous L*er* allele is always utilized as the denominator. The horizontal line indicates the 0.05 significance threshold. Note the different LOD scales on the chromosomes.

that *GS-Elong*, *GS-AOP*, and *EC198L* all control leaf aliphatic glucosinolate concentration via the same pathway or mechanism. As *GS-Elong* and *GS-AOP* are biosynthetic loci, this suggests that *EC198L* is also somehow altering glucosinolate biosynthesis.

QTL controlling total indolic glucosinolate accumulation: Indolic glucosinolates are produced from tryptophan whereas all aliphatic glucosinolates in Arabidopsis are produced from methionine (HALKIER and DU 1997). This fundamental difference suggests that their accumulation may be independently regulated. This theory is supported by the absence of a correlation between aliphatic and indolic glucosinolate levels in either leaves or seeds (leaf, $n = 162$, $r = 0.01$, $P = 0.19$; seed, $n = 162$, $r = 0.01$, $P = 0.87$). To further test this theory, we mapped QTL controlling the accumulation of total leaf and seed indolic glucosinolate concentration. This showed that there was no overlap between QTL controlling indolic and aliphatic glucosinolate accumulation (Figures 5 and 7). Thus, the two classes appear to be independently regulated in this cross.

The observation of a positive correlation between leaf and seed indolic glucosinolates suggested that some genes influence indolic glucosinolate levels in both tissues. Of the six identified QTL, only the *DF119L* QTL controls both leaf and seed indolic glucosinolate levels

TABLE 3

Interactions of QTL controlling leaf aliphatic glucosinolate concentration

Source	d.f.	Mean square	*F*-ratio	*P*
GS-AOP	1	284.5908	95.5034	0.0001
EC198L	1	123.3567	41.3962	0.0001
GS-ELONG	1	69.9811	23.4844	0.0001
GS-AOP × *GS-ELONG*	1	146.0271	49.0040	0.0001
EC198L × *GS-ELONG*	1	23.9201	8.0271	0.0053
GS-AOP × *EC198L*	1	4.6521	1.5611	0.2136
GS-AOP × *EC198L* × *GS-ELONG*	1	18.0367	6.0528	0.0151
Error	140	2.9799		

The variable tested was leaf aliphatic glucosinolate concentration. The 148 lines where all of the *GS-AOP*, *EC198L*, and *GS-Elong* markers had been scored were utilized. The $R^2 = 68\%$.

(Figure 7). The other five QTL appear to be tissue specific. As expected from the observation that L*er* contains significantly higher indolic glucosinolate levels in both tissues, the major QTL were such that the L*er* allele led to higher indolic glucosinolates. No epistasis was identified among any of the indolic QTL.

QTL for benzylic glucosinolates: The third glucosinolate class in Arabidopsis is the benzyl glucosinolates derived from phenylalanine, which are predominantly seed specific (Table 1). In comparison to the other two classes, very little is known about the quantitative or qualitative variation present in this class of glucosinolates. To investigate the genetic control of quantitative variation for the benzyl glucosinolates, we mapped QTL controlling the accumulation of benzyl glucosinolates in the seed. Strikingly, the major QTL controlling the accumulation of seed benzyl glucosinolate is linked to the *GS-Elong* locus that controls total leaf aliphatic glucosinolates (Figure 8). Our analysis of the RILs identified two different benzyl glucosinolates, which appear to differ in their carbon chain length (Kliebenstein *et al.* 2001b). Thus, it is possible that *GS-Elong* or a nearby locus controls the chain elongation of phenylalanine and this chain elongation leads to differences in accumulation. Further work is required to test if *GS-Elong*-controlled chain elongation is involved in benzyl glucosinolate biosynthesis. The other two QTL, near *GB120C* and *AD182C*, did not overlap with QTL from the other glucosinolate classes (Figure 8). Analysis of QTL interactions showed that there is a significant two-way interaction between *AD182C* and *GS-Elong*, suggesting that these loci control benzyl glucosinolate accumulation via the same mechanism (Table 4).

QTL for specific indolic glucosinolates: The variable used in mapping QTL for total leaf indolic glucosinolate concentration is the sum of the indolyl-3-methyl (I3M), 1-methoxy-indolyl-3-methyl (1MO-I3M), and 4-methoxy-indolyl-3-methyl (4MO-I3M) glucosinolates. However, seeds do not contain 4MO-I3M (Table 1). Mapping QTL controlling total indolic GS might obscure genetic segregation controlling the accumulation of specific glucosinolates. To test for this possibility we analyzed the individual indolic glucosinolate data for QTL and compared these QTL to those obtained using the composite trait.

This analysis showed that the use of summation variables concealed QTL controlling specific glucosinolates and obscured the effects of known QTL on specific glucosinolates (Figure 9). An example is the *FD167L* QTL that controls total leaf indolic glucosinolate concentration. Analyzing the specific glucosinolate concentrations showed that, while the *FD167L* L*er* allele increases total indolic glucosinolate levels, this allele only increases the level of I3M (Figure 9). In contrast the *FD167L* L*er* allele leads to decreased 4MO-I3M levels and has no effect on 1MO-I3M. This suggests that *FD167L* controls the ratio of the precursor I3M to the product 4MO-I3M and this affects the total level of indolic glucosinolates in the leaves.

In addition to hiding the effect of QTL influencing specific glucosinolates, the utilization of the summation variable for QTL mapping also hid glucosinolate-specific QTL. For example, a region at the top of chromosome I controls the accumulation of 1MO-I3M in the seed but has no effect on the total concentration of indolic glucosinolates in the seed. Similar QTL were found for 4MO-I3M in the leaf, while no I3M-specific QTL were identified. Thus, while the summation variable is a good tool to analyze total quantitative variation, it can be misleading when interpreting the quantitative variation of specific individual glucosinolates.

DISCUSSION

Genetics of glucosinolate quantitative variation in *A. thaliana*: Previous research into the genetics of quantitative variation in glucosinolates had investigated the accumulation of only one class of glucosinolates in one specific tissue type (Toroser *et al.* 1995; Uzunova *et al.* 1995). To better understand the control of glucosinolate quantitative variation we examined genetic control of the accumulation of three different glucosinolate

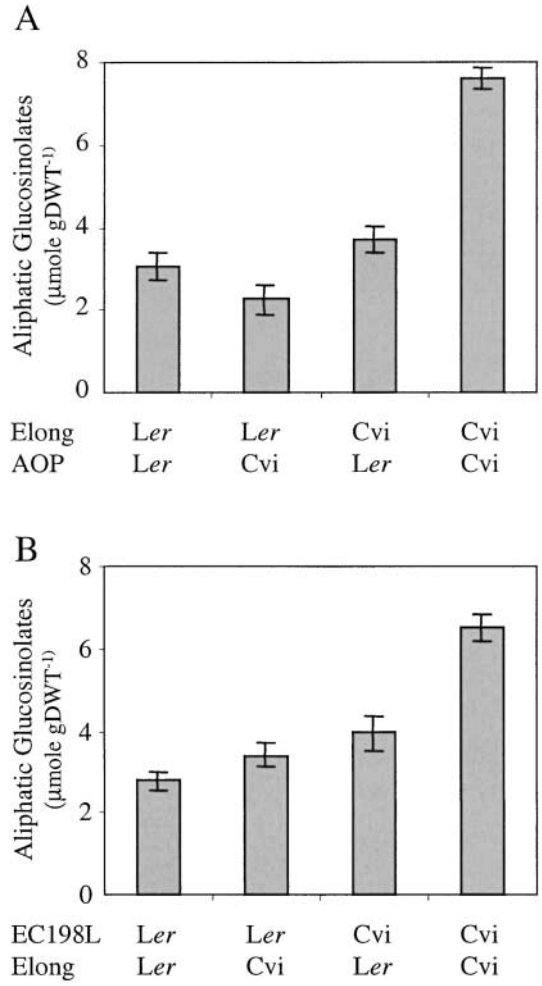

FIGURE 6.—Allele-specific interactions of QTL controlling leaf aliphatic glucosinolates. Bar diagrams show the average leaf aliphatic glucosinolate accumulation. Allelic states of the loci are listed below each bar. (A) Interaction of *GS-Elong* and *GS-AOP*. (B) Interaction of *GS-Elong* and *EC198L*.

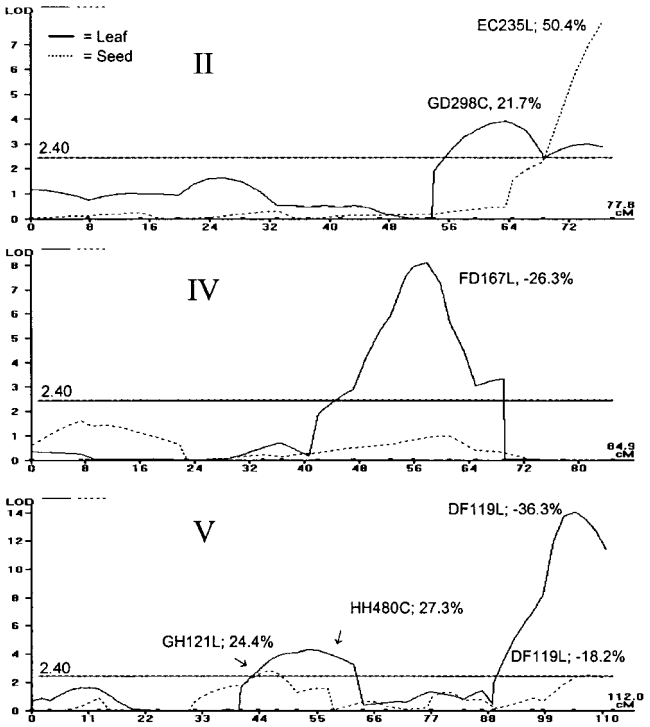

FIGURE 7.—QTL controlling indolic glucosinolate concentration in the leaves and seeds of the L*er* × Cvi RILs. LOD plots of QTL controlling the accumulation of total indolic glucosinolates are shown. Total indolic glucosinolate accumulation was determined by summing all of the individual tryptophan-derived glucosinolates. Numbers are the percentages of difference in mean glucosinolate concentration between lines homozygous for the Cvi allele and lines homozygous for the L*er* allele at the listed markers. The mean from the homozygous L*er* allele is always utilized as the denominator. The horizontal line indicates the 0.05 significance threshold. Note the different LOD scales on the chromosomes.

classes (aliphatic, indolic, and benzylic) in two different tissues (leaves and seeds) of *A. thaliana*. This analysis showed that there were a number of QTL controlling the accumulation of each of the three different classes (Figures 5, 7, and 8). Further, most of the QTL were specific for the accumulation of a specific glucosinolate class in a specific tissue type (Figures 5, 7, and 8). This large number of QTL suggests that it will be possible to utilize natural variation in *A. thaliana* to develop an in-depth understanding of glucosinolate biosynthesis.

Tissue specificity of QTL: The strong correlation between leaf and seed aliphatic glucosinolates suggests overlapping regulation between the two tissues. However, there was no overlap in QTL controlling aliphatic glucosinolate accumulation in these tissues (Figure 5). This suggests that there are no segregating global regulators of aliphatic glucosinolate concentration in this cross. It also suggests that the quantitative effects of *GS-AOP* and *GS-Elong* are tissue specific, whereas their qualitative effects are not tissue specific (Figure 5 and Table 1). The tissue specificity of *GS-AOP* could be explained by a difference in the aliphatic glucosinolates that accumulate in the two tissues. The predominant leaf aliphatic glucosinolates are the C_3 and C_4 alkenyl, hydroxyalkyl, or methylsulfinylalkyl glucosinolates, which are determined by the AOP and Elong enzymes (DE QUIROS *et al.* 2000; KLIEBENSTEIN *et al.* 2001a). In contrast, the predominant seed aliphatic glucosinolates are a mixture of C_7 and C_8 methylthioalkyl glucosinolates and C_3 and C_4 benzoyloxyalkyl glucosinolates. Thus, the impact of the individual glucosinolates derived from the AOP and Elong reactions would be reduced in the seeds in comparison to the leaves. This could explain the tissue specificity of these QTL.

Comparison of QTL for total glucosinolate levels and for individual glucosinolate levels: A number of glucosinolate QTL studies measure only total glucosinolate concentrations or sum together all the glucosinolates and then map only the composite trait (MITHEN and TOROSER 1995; TOROSER *et al.* 1995). These summation approaches have the possibility of hiding QTL for individ-

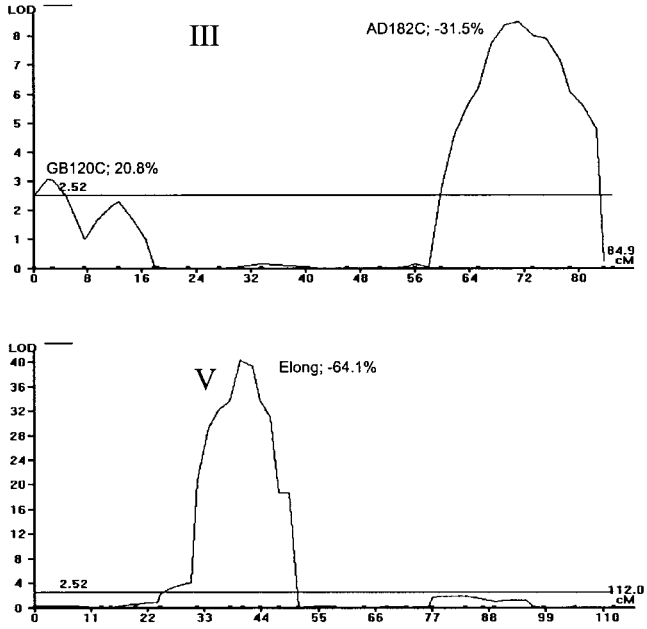

FIGURE 8.—QTL controlling seed benzyl glucosinolate concentration in the L*er* × Cvi RILs. LOD plots of QTL controlling the accumulation of total benzylic glucosinolates are shown. Total benzylic glucosinolate accumulation was determined by summing the values for both phenylalanine-derived glucosinolates. Numbers are the percentages of difference in mean glucosinolate concentration between lines homozygous for the Cvi allele and lines homozygous for the L*er* allele at the listed markers. The mean from the homozygous L*er* allele is always utilized as the denominator. The horizontal line indicates the 0.05 significance threshold. Note the different LOD scales on the chromosomes.

ual glucosinolates and hiding the impact of differential effects of the identified QTL on individual glucosinolates. To test for both of these possibilities, we compared the QTL controlling total indolic glucosinolate accumulation to QTL for individual glucosinolates. This showed that analysis of summed total glucosinolates obscured some QTL, which controlled the accumulation of specific glucosinolates. Further, some of the QTL that influenced the total accumulation of indolic glucosinolates actually had differential effects on the accumulation of individual glucosinolates. Thus, to be thorough, it is best to map QTL for all of the individual variables as well as total glucosinolate concentration.

Leaf aliphatic glucosinolates determine the base of seed aliphatic glucosinolate accumulation: Previous work has shown that glucosinolates can be transported from the leaves to the developing siliques (BRUDNELL *et al.* 1999). Further, siliques are also known to have an endogenous glucosinolate biosynthetic capacity (DU and HALKIER 1998). Thus, the concentration of seed glucosinolates may depend upon a combination of transport from the leaves and biosynthesis within the silique. If there were no variation in glucosinolate transport from the leaves to the seeds, the level of glucosinolates in the leaves would represent the minimal seed

TABLE 4

Interactions of QTL controlling seed benzylic glucosinolate concentration

Source	d.f.	Mean square	*F*-ratio	*P*
GS-ELONG	1	4.7597	290.3043	0.0001
AD182C	1	0.7389	45.0645	0.0001
GB120C	1	0.1759	10.7291	0.0013
AD182C × *GS-ELONG*	1	0.1163	7.0959	0.0086
AD182C × *GB120C*	1	0.0140	0.8563	0.3563
GS-ELONG × *GB120C*	1	0.0020	0.1222	0.7272
Error	149	0.0164		

The variable tested was seed benzylic glucosinolate concentration. The data from 156 lines where all of the *GS-Elong*, *AD182C*, and *GB120C* markers had been scored were utilized and the $R^2 = 74\%$.

glucosinolate concentration. This level could then be enhanced by *de novo* biosynthesis in the silique. Analysis of leaf and seed aliphatic glucosinolate accumulation in the 162 L*er* × Cvi RILs showed a significant positive correlation (Figure 4). Interestingly, there appears to be a threshold effect to this correlation, with the aliphatic leaf glucosinolates representing a basal concentration of aliphatic seed glucosinolates (Figure 4). Further, a study investigating the accumulation of aliphatic leaf and seed glucosinolates in 35 Arabidopsis ecotypes showed a similar positive correlation, with aliphatic leaf glucosinolates acting as a lower boundary on seed glucosinolates (KLIEBENSTEIN *et al.* 2001b).

Side-chain modification and regulation of glucosinolate concentration: The major QTL controlling the accumulation of leaf glucosinolates in the L*er* × Cvi RILs was tightly linked to the previously cloned glucosinolate biosynthetic locus, *GS-AOP* (Figure 5, Table 3). This locus is the result of a tandem gene duplication where differential expression of one gene produces alkenyl glucosinolates (AOP2) or the other gene generates hydroxyalkyl glucosinolates (AOP3; KLIEBENSTEIN *et al.* 2001a). L*er* expresses the AOP3 gene and produces hydroxypropyl glucosinolate while Cvi expresses the AOP2 gene to produce alkenyl glucosinolates (Figures 2 and 3). The enzymes encoded by these genes function to modify the side chains of aliphatic glucosinolates at the end of the glucosinolate biosynthetic pathway (Figure 1). These data suggest that alternative *GS-AOP* haplotypes cause large differences in leaf aliphatic glucosinolate concentration. One possible explanation is that the *GS-AOP* locus controls glucosinolate biosynthesis by regulating metabolic flux through the pathway and that the enzyme encoded by AOP2 has a higher enzymatic rate than that encoded by AOP3. An alternative explanation is that hydroxyl glucosinolates may be more effective feedback inhibitors than alkenyl glucosinolates. We cannot currently differentiate between these alterna-

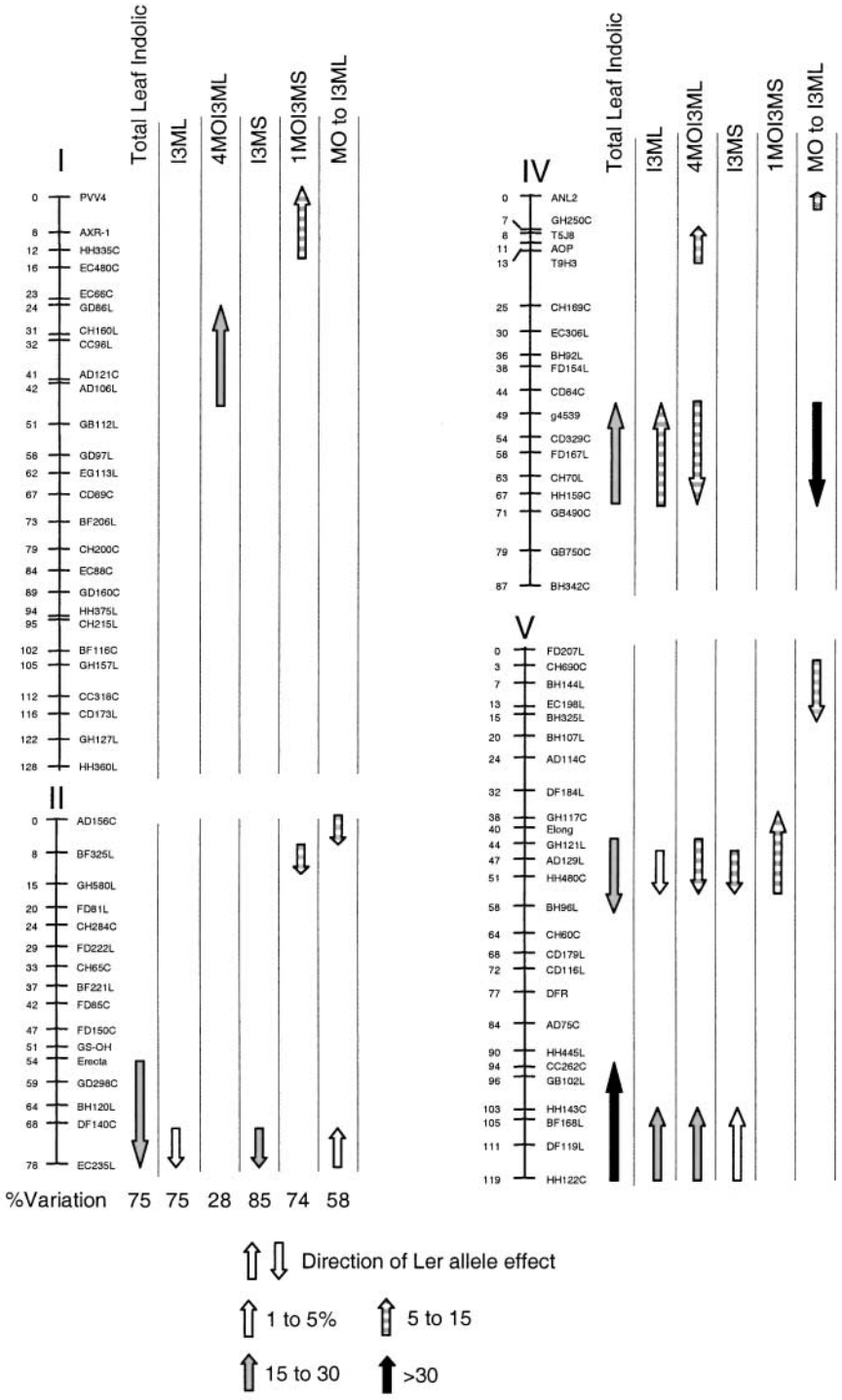

FIGURE 9.—Compound specificity of indolic glucosinolate QTL. Diagrams show the location and strength of QTL for the traits denoted by the columns. The QTL are shown as arrows to the right of the chromosomal map. The total percentage of the variation explained by the QTL for the given trait is listed at the bottom of each column. Chromosome III is not shown as it did not contain any QTL. The arrow height represents the range of 0.05 significance for the QTL. S represents seed glucosinolates and L represents leaf glucosinolates. MO to I3M = (1-methoxy-I3M + 4-methoxy-I3M)/I3M.

tives. Finally, the mapping data reported here cannot exclude the possibility that another tightly linked locus is responsible for the variation in glucosinolate accumulation.

Effect of methionine elongation on aliphatic glucosinolate accumulation: Arabidopsis contains aliphatic glucosinolates with differing carbon lengths, which are produced by a common elongation pathway (HOGGE *et al.* 1988; DE QUIROS *et al.* 2000; KLIEBENSTEIN *et al.* 2001b). One locus, *GS-Elong*, has been identified that controls C_3 *vs.* C_4 polymorphism (DE QUIROS *et al.* 2000). Previous work had shown that this locus was a major determinant of glucosinolate concentration in seeds from the L*er* × Col RILs of Arabidopsis (MITHEN and CAMPOS 1996). In our analysis of the L*er* × Cvi RILs, the *GS-Elong* locus exerted a significant influence on leaf but not on seed aliphatic glucosinolate concentrations (Figure 5). This suggests that the quantitative effect of *GS-Elong* is either dependent upon the genetic background in which it occurs or that the C_4 alleles present in Cvi and Col

are somehow different. The hypothesis that *GS-Elong* is dependent upon the genetic background is supported by the observation that there is an epistatic interaction between *GS-Elong* and both *GS-AOP* and *EC198L* (Figure 6). Further work is required to understand how the C_3/C_4 biochemical polymorphism controls total accumulation of aliphatic glucosinolates.

Independent regulation of glucosinolates derived from different amino acids: The accumulation of glucosinolates derived from different amino acids could be coregulated because they utilize the same enzymatic reactions in the formation of their basic glucosinolate skeleton (Figure 1; Halkier and Du 1997). However, in the L*er* × Cvi RILs there is no correlation between the production of indolic, aliphatic, and benzylic glucosinolates. This agrees with previous analysis of Arabidopsis ecotypes, which also showed a lack of correlation between glucosinolates produced from different amino acids (Kliebenstein *et al.* 2001b). Further, only one QTL, *GS-Elong*, appears to coregulate the accumulation of two different glucosinolates and it is unknown if this is due to the action of a single locus or closely linked loci (Figures 5, 7, and 8). This suggests that the enzymes involved in biosynthesis are either amino acid specific or do not regulate glucosinolate accumulation. The cytochrome P450 monooxygenases are known to have specificity for certain amino acids and could thereby explain some of this independent regulation (Kahn *et al.* 1999; Wittstock and Halkier 2000). In contrast, sulfotransferase and glucosyltransferase have minimal side-chain specificity and might thereby have been expected to lead to glucosinolate coregulation (Halkier and Du 1997). However, this is not the case in this study.

Our study shows that glucosinolate concentration and type are regulated by a large number of naturally varying genetic loci. This information provides precise genetic tools to manipulate both glucosinolate type and amount and test their relative biological effects. Further, the observation that glucosinolate type and concentration are highly variable in Arabidopsis suggests that breakdown type and amount may be variable as well. If this is the case, then it may be possible to utilize Arabidopsis genetics to compare the relative impacts of glucosinolate type and amount and breakdown product type and amount upon plant fitness in a single species.

We acknowledge Antje Figuth for technical support in extracting all glucosinolates, Deana Pedersen for aid in plant growth and tissue collection, Michael Reichelt for assistance in compound identification, and Jonathan Gershenzon for editorial advice. This work was funded by the Max-Planck Gesellschaft and grant DEB-9527725 from the U.S. National Science Foundation.

LITERATURE CITED

Alonso-Blanco, C., A. J. M. Peeters, M. Koornneef, C. Lister, C. Dean *et al.*, 1998 Development of an Aflp based linkage map of Ler, Col and Cvi Arabidopsis thaliana ecotypes and construction of a Ler/Cvi recombinant inbred line population. Plant J. **14:** 259–271.

Alonso-Blanco, C., H. Blankestijn-de Vries, C. J. Hanhart and M. Koornneef, 1999 Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **96:** 4710–4717.

Basten, C. J., B. S. Weir and Z-B. Zeng, 1999 *QTL Cartographer, Version 1.13*. Department of Statistics, North Carolina State University, Raleigh, NC.

Bones, A. M., and J. T. Rossiter, 1996 The myrosinase-glucosinolate system, its organisation and biochemistry. Physiol. Plant. **97:** 194–208.

Brudnell, A. J. P., H. Griffiths, J. T. Rossiter and D. A. Baker, 1999 The phloem mobility of glucosinolates. J. Exp. Bot. **50:** 745–756.

de Quiros, H. C., R. Magrath, D. McCallum, J. Kroymann, D. Schnabelrauch *et al.*, 2000 α-Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. Theor. Appl. Genet. **101:** 429–437.

Donkin, S. G., M. A. Eiteman and P. L. Williams, 1995 Toxicity of glucosinolates and their enzymatic decomposition products to Caenorhabditis Elegans. J. Nematol. **27:** 258–262.

Du, L., and B. Halkier, 1998 Biosynthesis of Glucosinolates in the developing silique walls and seeds of sinapis alba. Phytochemistry **48:** 1145–1150.

Graser, G., B. Schneider, N. Oldham and J. Gershenzon, 2000 The methionine chain elongation pathway in the biosynthesis of glucosinolates in *Eruca sativa* (Brassicaceae). Arch. Biochem. Biophys. **378:** 411–419.

Halkier, B. A., and L. Du, 1997 The biosynthesis of glucosinolates. Trends Plant Sci. **2:** 425–431.

Hogge, L. R., D. W. Reed, E. W. Underhill and G. W. Haughn, 1988 HPLC separation of glucosinolates from leaves and seeds of *Arabidopsis thaliana* and their identification using thermospray liquid chromatography-mass spectometry. J. Chrom. Sci. **26:** 551–556.

Kahn, R. A., T. Fahrendorf, B. A. Halkier and B. L. Moller, 1999 Substrate specificity of the cytochrome P450 enzymes CYP79A1 and CYP71E1 involved in the biosynthesis of the cyanogenic glucoside dhurrin in Sorghum bicolor (L.) Moench. Arch. Biochem. Biophys. **363:** 9–18.

Kliebenstein, D., V. Lambrix, M. Reichelt, J. Gershenzon and T. Mitchell-Olds, 2001a Gene duplication and the diversification of secondary metabolism: side chain modification of glucosinolates in *Arabidopsis thaliana*. Plant Cell **13:** 681–693.

Kliebenstein, D. J., J. Kroymann, P. Brown, A. Figuth, D. Pedersen *et al.*, 2001b Genetic control of natural variation in *Arabidopsis thaliana* glucosinolate accumulation. Plant Physiol. **126:** 811–825.

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 natural populations. Genomics **1:** 174–181.

Li, Q., S. Eigenbrode, G. Stringham and M. Thiagarajah, 2000 Feeding and growth of *Plutella xylostella* and *Spodoptera eridania* on *Brassica juncea* with varying glucosinolate concentrations and myrosinase activities. J. Chem. Ecol. **26:** 2401–2419.

Manici, L., L. Lazzeri and S. Palmieri, 1997 In vitro fungitoxic activity of some glucosinolates and their enzyme-derived products toward plant pathogenic fungi. J. Agric. Food Chem. **45:** 2768–2773.

Mithen, R., and H. Campos, 1996 Genetic variation of aliphatic glucosinolates in *Arabidopsis thaliana* and prospects for map based gene cloning. Entomol. Exp. Appl. **80:** 202–205.

Mithen, R., and D. Toroser, 1995 Biochemical genetics of aliphatic glucosinolates in Brassica and Arabidopsis. Soc. Exp. Biol. Semin. Ser. **56:** 261–275.

Mithen, R. F., B. G. Lewis and G. R. Fenwick, 1986 *In vitro* activity of glucosinolates and their products against *Leptosphaeria maculans*. Trans. Br. Mycol. Soc. **87:** 433–440.

Mithen, R., J. Clarke, C. Lister and C. Dean, 1995 Genetics of aliphatic glucosinolates.3. Side chain structure of aliphatic glucosinolates in Arabidopsis thaliana. Heredity **74:** 210–215.

Rossiter, J. T., D. C. James and N. Atkins, 1990 Biosynthesis of 2-hydroxy-3-butenylglucosinolate and 3-butenylglucosinolate in *Brassica napus*. Phytochemistry **29:** 2509–2512.

Siemens, D. H., and T. Mitchell-Olds, 1996 Glucosinolates and

herbivory by specialist herbivores: consequences of concentration and induced resistance. Environ. Entomol. **80:** 231–237.

Stowe, K. A., 1998 Realized defense of artificially selected lines of Brassica rapa—Effects of quantitative genetic variation in foliar glucosinolate concentration. Environ. Entomol. **27:** 1166–1174.

Toroser, D., C. Thormann, T. Osborn and R. Mithen, 1995 RFLP mapping of quantitative trait loci controlling seed aliphatic glucosinolate content in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. **91:** 802–808.

Uzunova, M., W. Ecke, K. Weissleder and G. Robbelen, 1995 Mapping the genome of rapeseed (Brassica napus L.).1. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor. Appl. Genet. **90:** 194–204.

Verhoeven, D., H. Verhagen, R. Goldbohm, P. van den Brandt and G. van Poppel, 1997 A review of mechanisms underlying anticarcinogenicity by Brassica vegetables. Chem.-Biol. Interact. **103:** 79–129.

Wink, M., 1988 Plant Breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. Theor. Appl. Genet. **75:** 225–233.

Wittstock, U., and B. A. Halkier, 2000 Cytochrome P450CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of the benzylglucosinolate. J. Biol. Chem. **275:** 14659–14666.

Communicating editor: V. Sundaresan