DOI: 10.1534/genetics.106.056986

Natural Variation in *MAM* Within and Between Populations of *Arabidopsis lyrata* Determines Glucosinolate Phenotype

Andrew J. Heidel,*[,1] Maria J. Clauss,* Juergen Kroymann,* Outi Savolainen[†] and Thomas Mitchell-Olds*[,2]

*Department of Genetics and Evolution, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany and [†]Department of Biology, University of Oulu, FIN-90014 Oulu, Finland

Manuscript received February 13, 2006
Accepted for publication May 3, 2006

ABSTRACT

The genetic variation that underlies the glucosinolate phenotype of *Arabidopsis lyrata* ssp. *petraea* was investigated between and within populations. A candidate glucosinolate biosynthetic locus (*MAM*, containing methylthioalkylmalate synthase genes) was mapped in *A. lyrata* to a location on linkage group 6 corresponding to the homologous location for *MAM* in *A. thaliana*. In *A. thaliana MAM* is responsible for side chain elongation in aliphatic glucosinolates, and the *MAM* phenotype can be characterized by the ratios of long- to short-chain glucosinolates. A quantitative trait loci (QTL) analysis of glucosinolate ratios in an *A. lyrata* interpopulation cross found one QTL at *MAM*. Additional QTL were identified for total indolic glucosinolates and for the ratio of aliphatic to indolic glucosinolates. *MAM* was then used as the candidate gene for a within-population cosegregation analysis in a natural *A. lyrata* population from Germany. Extensive variation in microsatellite markers at *MAM* was found and this variation cosegregated with the same glucosinolate ratios as in the QTL study. The combined results indicate that both between- and within-population genetic variation in the *MAM* region determines phenotypic variation in glucosinolate side chains in *A. lyrata*.

DETERMINATION of the genetic variation that underlies ecologically important phenotypic variation is a central goal in functional genomics. To identify genetic variation a number of powerful tools are available, including quantitative trait loci (QTL) analysis and candidate gene studies. These methods are useful in the identification of genes for future functional studies. QTL studies can estimate the number of variable loci affecting a trait (reviewed by Koornneef *et al.* 2004). However, most QTL experiments compare only two inbred genomes at a time from widely separated populations rather than the functional variants segregating within natural populations. While multiple QTL crosses can be studied for a trait of interest (Symonds *et al.* 2005), such studies are time consuming and still characterize only a subset of allelic diversity, especially in polymorphic, outcrossing species. Yet this uncharacterized allelic diversity may be critical to understanding the evolution of complex traits.

Alternatively, studies of candidate genes can determine whether variability at an individual locus influences phenotypic variability. These studies analyze either genotype–phenotype associations within a population or cosegregation of phenotypes with molecular markers within families, using polymorphisms in or near functionally important loci. This approach is feasible only for known genes and pathways that influence well-studied traits. Population genetic analysis of variation at these genes in natural populations can reveal additional information about evolutionary processes and history. To overcome the limited focus on a small number of candidate genes, it has been suggested that a combination of QTL and association studies are required to understand genetic architecture (Barton and Keightley 2002). Here we combine genomewide QTL mapping between populations and candidate gene cosegregation analysis within a population to infer the genes controlling natural variation in secondary metabolism between and within natural populations.

Glucosinolates are secondary metabolites that are found in the Brassicaceae and contribute to defense against herbivores and pathogens. Glucosinolates affect insect herbivores (Lambrix *et al.* 2001; Kroymann *et al.* 2003; Kliebenstein *et al.* 2005) and exhibit extensive structural variation both within and between species (Kliebenstein *et al.* 2001b; Windsor *et al.* 2005). Significant heritability for glucosinolates has also been observed within a single undisturbed natural population of *Arabidopsis lyrata* (Clauss *et al.* 2006). Aliphatic and indolic glucosinolates are derived from aliphatic (methionine, alanine, valine, leucine, and isoleucine)

[1]*Corresponding author:* Fritz Lipmann Institute, Beutenbergstrasse 11, 07745 Jena, Germany. E-mail: aheidel@ice.mpg.de

[2]*Present address:* Department of Biology, Duke University, Durham, NC 27708.

and indolic amino acids (tryptophan), respectively (Wittstock and Halkier 2002). In *A. thaliana* and *A. lyrata*, aliphatic glucosinolates are formed exclusively from methionine (Windsor *et al.* 2005). In addition to a group of enzymes encoded at the *MAM* locus, biosynthesis of methionine-derived glucosinolates is controlled by enzymes of the CYP79 family (convert amino acids to aldoxime) and of the CYP83 family (perform the subsequent step in core structure synthesis) (Kliebenstein *et al.* 2005; Windsor *et al.* 2005). Biosynthesis of indolic glucosinolates is also controlled by CYP79 and CYP83 family enzymes (Mikkelsen *et al.* 2000; Windsor *et al.* 2005).

In *A. thaliana*, the *MAM* locus comprises a small, tandemly linked gene family on chromosome 5, encoding enzymes that catalyze side-chain elongation of methionine precursors of aliphatic glucosinolates (Kroymann *et al.* 2001, 2003). Two unlinked genes on chromosome 1 in *A. thaliana* share sequence similarity with *MAM* locus genes and one encodes isopropylmalate synthase (IPMS) activity (Kroymann *et al.* 2001; Field *et al.* 2004). Different haplotypes at the complex *MAM* locus (also referred to as *GS-Elong*) alter the predominant side-chain lengths of aliphatic glucosinolates in *A. thaliana*. Allelic variation between the *MAM* alleles of *A. thaliana* lines Col and CL5 causes a QTL for glucosinolates and resistance to the generalist herbivore *Spodoptera exigua* (Kroymann *et al.* 2003). Likewise, the *MAM* QTL between Ler-0 and Cape Verde Islands (Cvi-0) also influences glucosinolates (Kliebenstein *et al.* 2001a). A QTL spanning the *MAM* locus influenced resistance to the Brassicaceae specialist *Plutella xylostella* in the Ler-0 × Cvi-0 cross (Kliebenstein *et al.* 2002). The *MAM* region was not directly detected as a QTL for *Trichoplusia ni* resistance in the same cross, but showed a significant interaction with another QTL at the *GS-AOP* glucosinolate biosynthetic locus (Kliebenstein *et al.* 2002). Therefore, in *A. thaliana*, *MAM* has demonstrated genetic variation that determines glucosinolate phenotypic variation, which in turn affects herbivory resistance.

These QTL studies have unraveled some of the genetic causes of glucosinolate variation between *A. thaliana* accessions. We turn our attention to a closely related species, *A. lyrata*, to test the generality of this functional variation. Furthermore, in *A. lyrata* we can explore the causes and consequences of variation segregating within populations because this self-incompatible perennial herb maintains high within-population diversity at neutral and functional loci (Clauss and Mitchell-Olds 2006). *A. lyrata* exhibits extensive glucosinolate variation between (Windsor *et al.* 2005) and within populations (Clauss *et al.* 2006) and, in particular, significant heritability for the ratios of short side- to long side-chain glucosinolates (Clauss *et al.* 2006). We first examine genetic variation affecting glucosinolates between two European populations of *A. lyrata* using genomewide QTL mapping. The QTL analysis allows us to test whether *MAM* affects the glucosinolate phenotype and identify other loci that also affect glucosinolate variation. Next we focus on *MAM*, using it in a candidate gene analysis to infer whether within-population genetic variation at *MAM* influences glucosinolate variability. This combined approach can identify whether *MAM* determines the glucosinolate phenotype in *A. lyrata* as it does in *A. thaliana* and illustrates how functional variability in *MAM* influences the glucosinolate phenotype among and within populations.

MATERIALS AND METHODS

Study species and plant: *A. lyrata* ssp. *petraea* (henceforth *A. lyrata*) is a self-incompatible perennial herb found in temperate Eurasia. For the cosegregation analysis we studied a large diploid population of *A. lyrata* near Plech, Bavaria, Germany (49°55′N; 11°31′E) (Koch *et al.* 2001). The Plech population is physically isolated from adjacent *A. lyrata* localities, with individual plants distributed discontinuously on dolomitic rock outcrops (Clauss *et al.* 2002). Microsatellite markers and sequence polymorphisms in functional genes indicate random mating and high levels of diversity within the Plech population (Clauss *et al.* 2002; Clauss and Mitchell-Olds 2003). Unlike the ephemeral, inbreeding life history of *A. thaliana*, the outcrossing mating system and perennial growth habit of *A. lyrata* are representative of many plants in undisturbed temperate habitats (Schierup 1998; Karkkainen *et al.* 1999; Mitchell-Olds 2001; Clauss *et al.* 2002).

Karhumäki × Mjällom cross: As described by Kuittinen *et al.* (2004), two *A. lyrata* individuals from a Mjällom (Sweden) population and two individuals from a Karhumäki (Russia) population were crossed (K × M cross). One Mjällom plant was crossed to one Karhumäki plant while the other Mjällom plant was crossed to the other Karhumäki plant. Resulting F_1 progeny were reciprocally crossed to produce F_2 progeny and 94 of these F_2 genotypes were used for glucosinolate and QTL analysis. For each genotype, six replicates were reproduced clonally via tissue culture. Individual plants from each genotype were grown together in groups and genotype means were analyzed for QTL mapping. The plants were grown in 7 × 7 × 7.5 cm pots in a controlled environment growth chamber with a 13-hr day and 19°/16° day/night temperature. At the time of sampling for glucosinolates, mortality had reduced the number of individuals, but all genotypes had at least three living clones and 92 of 94 genotypes had at least four clones. Plants were sampled for DNA after sampling for glucosinolate analysis. Three of the lines did not amplify at the *MAM* markers, leaving 91 lines that were genotyped at the *MAM* markers.

Plech crosses: The crosses used are fully described by Clauss *et al.* (2006). Unique seed families were collected from the field and one individual per family was grown to reproductive maturity in a growth chamber and randomly assigned to a pair. Fifteen pairs were crossed. We analyzed the glucosinolate and genetic markers of the full sibs resulting from these 15 crosses. At the point of sampling for DNA, mortality had reduced the number of plants to a minimum of 9 plants in one family and a maximum of 18 in another family, with a mean of 13.1 plants per family sampled. Differences among sibships appear as the FAMILY effect in ANOVA and are unrelated to the genotype–phenotype cosegregation, which is the goal of this analysis.

Glucosinolate extraction and analysis: *K × M cross:* Leaf tissue was collected from each individual plant separately 7 months after clones were transferred from agar to soil, weighed (~0.05 g), and frozen in liquid N_2. Glucosinolates

were extracted as described (Kliebenstein *et al.* 2001c) with two modifications in the first extraction step with methanol: (1) 120 μl of water was used instead of 80 μl and (2) sinalbin was added as an internal standard.

For HPLC analysis 60 μl of the sample was injected. HPLC conditions were as described (Kliebenstein *et al.* 2001c) except that the following program was used: an 8-min gradient from 1.5 to 5% acetonitrile, a 2-min gradient to 7% acetonitrile, a 25-min gradient to 42% acetonitrile, a 1-min gradient to 92%, 1 min at 92% acetonitrile, a 1-min gradient to 2% acetonitrile, and a final 1-min gradient to 1.5% acetonitrile.

Glucosinolates were identified by retention time and absorption spectrum. The following glucosinolates were found: (1) 3-methylsulfinylpropyl, (2) 6-methylsulfinylhexyl, (3) 3-methylthiopropyl, (4) 7-methylsulfinylheptyl, (5) 8-methylsulfinyloctyl, (6) indolyl-3-methyl, (7) 4-methoxy-indolyl-3-methyl, (8) 1-methoxy-indolyl-3-methyl, (9) 7-methylthioheptyl. No hydroxy glucosinolates were found.

Each glucosinolate peak was standardized first by internal glucosinolate standard and then by wet weight. To calculate the total aliphatic and indolic glucosinolate values for QTL analysis, the amount of each individual aliphatic (glucosinolates 1, 2, 3, 4, 5, and 9 above) and indolic (glucosinolates 6, 7, and 8 above) glucosinolate was added together respectively. Except for the ratio of aliphatic to indolic glucosinolates all subsequent glucosinolate data analysis was done within the categories of aliphatic and indolic glucosinolates. The aliphatic and indolic glucosinolates detected here each have nondistinguishable response factors within the margin of error in their own group. Therefore, all data, except for the ratio of total aliphatic to total indolic glucosinolates, were not corrected for response factors. The total aliphatic to total indolic glucosinolates ratio was corrected for response factors (Brown *et al.* 2003) before other standardization. The totals of individual glucosinolates for different clones of the same genotype were then averaged and multiplied by a scalar to produce a standard deviation (2.51) suitable for the QTL software.

To calculate the aliphatic glucosinolate ratios for QTL analysis, the aliphatic glucosinolates were classified into four categories on the basis of the side-chain length: C3 for glucosinolates with a three-carbon side chain, C6 for a six-carbon side chain, C7 for a seven-carbon side chain, and C8 for an eight-carbon side chain. Ratios comparing the amount of glucosinolates with different chain lengths were then calculated: R3 = (C6 + C7 + C8)/C3, R6 = (C7 + C8)/C6, and R7 = C8/C7. The ratios were then averaged and multiplied by a scalar as described above. The inverse of R3 was used in all analysis and figures with the K × M cross due to low concentrations of C3 glucosinolates. These ratios reflect the activity of *MAM* locus enzymes by comparing substrate to products.

Plech population: Leaves were harvested at 8 weeks after germination. The extraction protocol was identical to that used with K × M cross except that samples were ~0.10 g and 40 μl of the sample was injected. The aliphatic glucosinolate ratios were corrected for response factors.

Genotyping: For both the K × M and Plech crosses, DNA was extracted using a scaled-down modification of Dellaporta *et al.* (1985). The CoS1 and CoS3 microsatellites were identified during partial shotgun sequencing of a BAC from *A. lyrata* ssp. *lyrata* (generously provided by June Nasrallah). These microsatellites flank a small gene family of *MAM* genes in this region (J. Kroymann, unpublished data). The microsatellites were amplified using the following primers:

CoS1f: 5′-GGG TCA TTA GCT CTG AAG TTG-3′
CoS1r: 5′-GTA CTA CTT TGA TGG ACA CTT G-3′
CoS3f: 5′-CAT GCA GTC TTA TAT ATT CTG ATG-3′
CoS3r: 5′-GTG ATG TGC TTT TGG AGT AAT TG-3′

The forward primers of CoS1 and CoS3 were 5′ labeled with Hex and Rox, respectively. Amplification conditions were standard ($MgCl_2$ 1.5 mm, annealing temperature: CoS1 and CoS3, 52.4°). PCR product length was determined by Trait-Genetics (Gatersleben, Germany) using capillary columns.

There is some ambiguity regarding *MAM* annotations; Field *et al.* (2004) refer to the *MAM1* and *MAM-L* genes as At5g23020 and At5g23010, respectively, while Kroymann *et al.* (2003) refers to *MAM1* and *MAM-L* genes as At5g23010 and At5g23020, respectively. The classification used by Kroymann *et al.* (2003) is accepted by TAIR and GenBank. Furthermore, there are some questions regarding what genes are included in the *MAM* family, leading to ambiguity in the definition of the *MAM* locus. Field *et al.* (2004) regard the IPMS genes At1g18500 and At1g74040 (located on a different chromosome than *MAM1* and *MAM-L* in *A. thaliana*) as members of the *MAM* family, annotating them *MAML-4* and *MAML-3*, respectively. While there is some homology between the *MAML-4* and *MAML-3* genes and the *MAM1*, *MAM2*, and *MAM-L* genes, indicating a common origin, cluster analysis of the genes (Kroymann *et al.* 2001) demonstrates that At1g18500 and At1g74040 are more closely related to IPMS genes in tomato than to *MAM1* or *MAM-L*. Due to an evolutionary distance of >40 million years, At1g18500 and At1g74040 are not considered part of the *MAM* family in this study. Therefore, in this study the *MAM* locus refers to only the locus containing *MAM1*, *MAM2*, and *MAM-L* as defined by Kroymann *et al.* (2001, 2003) or the corresponding region in *A. lyrata*.

Statistical analysis: QTL mapping was performed using Multimapper/outbred software (Sillanpää and Arjas 1999). This software performs Bayesian QTL mapping of outbred offspring and allows the use of parental genetic data as well as controlling for background variation using marker covariates. The genetic data used are described by Kuittinen *et al.* (2004) with the addition of the CoS markers. The priorlimit file used the following parameters for all analyses: line 1, [−13, 13]; line 2, [0, 2.51]; line 3, [0, 100]; line 4, [−13, 13]; line 5, [0, 100]. The parameters for the randomwalk files are given in supplemental Table 1 at http://www.genetics.org/supplemental/.

For the candidate gene analysis, the effects of FAMILY and *MAM* on glucosinolate ratios were determined by ANOVA with MAM and FAMILY as fixed effects and *MAM* nested within FAMILY. *MAM* was nested within FAMILY since alleles segregate within families. The glucosinolate ratios R3 and R7 were log transformed to conform to normality requirements, while the glucosinolate ratio R6 did not require transformation. The R3 ratio was also log transformed in the ANOVAs comparing *MAM* markers and *nga249*. All statistical analyses were performed using JMP 4 (SAS Institute).

RESULTS

Location of *MAM*: To locate *MAM* within the *A. lyrata* genome the segregation patterns of the two *MAM* markers CoS1 and CoS3 were compared to the patterns of the other markers in the K × M cross (Kuittinen *et al.*, 2004). No recombination was observed between the two *MAM* markers in the 94 individuals sampled, so the two markers were treated as one marker in further analyses. Given the anticipated short distance of <1 cM between the markers, the lack of recombination between the markers is expected. The *MAM* markers were found to be linked to markers on the sixth linkage group. The *MAM* markers were mapped within the

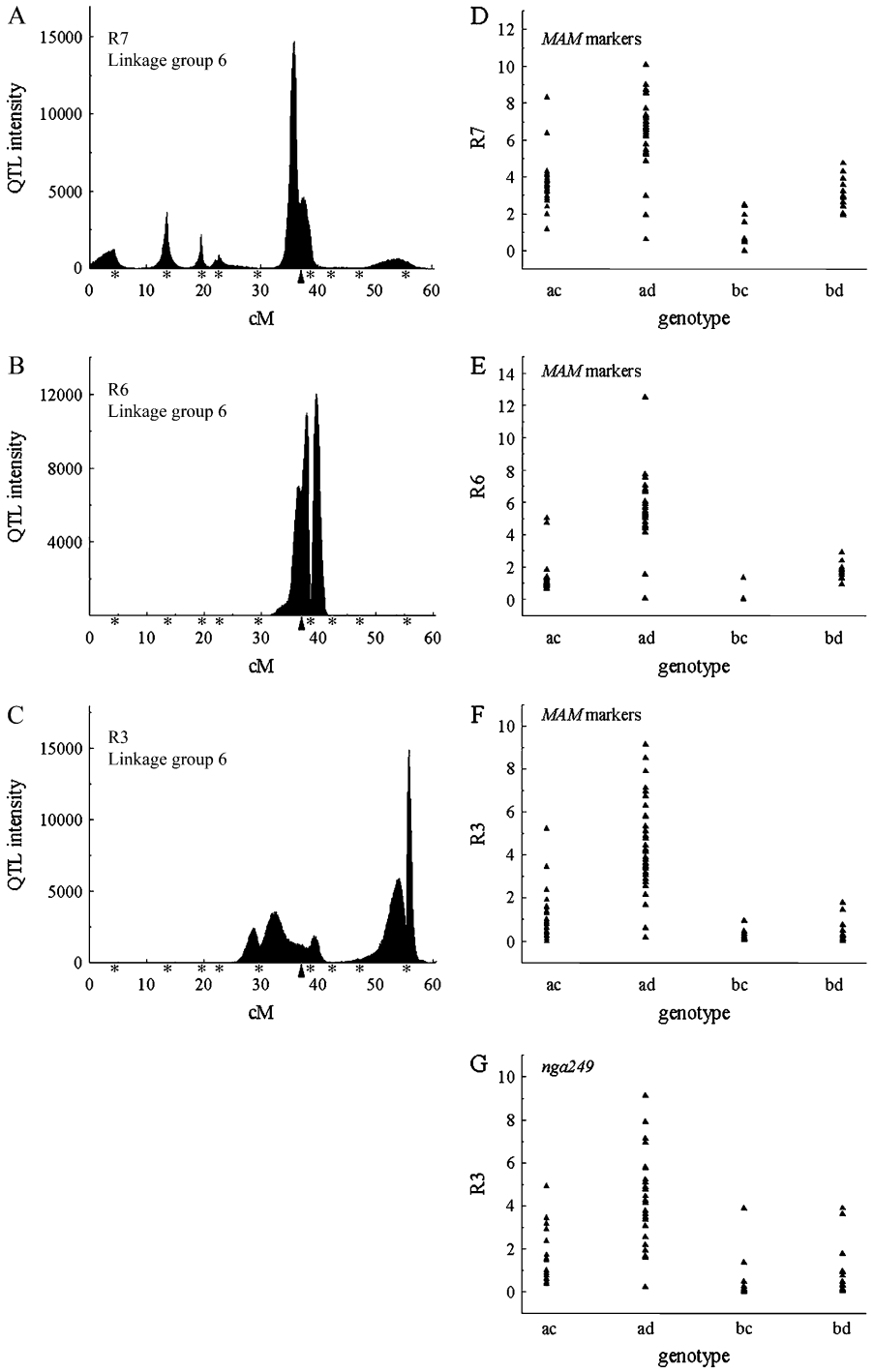

FIGURE 1.—(A–C) Bayesian posterior QTL intensity graphs for the three aliphatic glucosinolate ratios: R7, R6, and R3. Only linkage groups with QTL are displayed. The position of genetic markers is indicated below the *x*-axis with asterisks except for the *MAM* markers. The *MAM* markers are at 37 cM on linkage group 6 and are indicated with an arrowhead. (D–F) Glucosinolate ratios for the four *MAM* genotypes. (G) R3 glucosinolate ratio at *nga249*. The allele names to describe *nga249* are not the labels used by KUITTINEN *et al.* (2004). Instead the names are switched as needed to keep the labeling of haplotypes consistent with the *MAM* haplotypes.

linkage group by standard recombination mapping to a location 7.5 cM below *CLC-d* and 1.7 cM above *AthCDPK9* or 37 cM below the first marker on the linkage group. This location agrees with the arrangement of the *A. lyrata* linkage groups relative to *A. thaliana* as found by KUITTINEN *et al.* (2004).

There is significant transmission ratio distortion (segregation distortion) for the *MAM* markers (Pearson statistic $P = 0.0003$). This result is consistent with the transmission ratio distortion previously found on both flanking markers (KUITTINEN *et al.* 2004). Although the cause of the transmission ratio distortion in the *MAM* region is unknown, there is little reason to believe that *MAM* itself might be responsible.

QTL for glucosinolates: After adding the *MAM* markers to the *A. lyrata* map data from KUITTINEN *et al.* (2004), QTL were mapped for the three aliphatic glucosinolate ratios. These ratios of long side- to short side-chain aliphatic glucosinolates have been shown to be affected by different *MAM* haplotypes in *A. thaliana*, a

TABLE 1

Interpopulation differences linked to *MAM* and *nga249* influencing glucosinolate ratio R3

	R3		
Source	d.f.	SS	F
MAM markers	3	20.4	47.3****
nga249	3	15.0	22.4****

Fixed effects ANOVA comparing effects of linked markers at *MAM* and *nga249* on the R3 ratio of long/short glucosinolate concentrations segregating in the K × M cross. SS, sum of squares. *Significant *F*-values at $P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

logical result given that *MAM* enzymes lengthen the side chains of aliphatic glucosinolates (Kroymann *et al.* 2001; Textor *et al.* 2004). Two ratios had only one QTL, located at *MAM* for R7 and R6 (Figure 1, A and B). The R3 ratio had two QTL peaks, one at *MAM* and one 20 cM below *MAM* (Figure 1C). The four different *MAM* genotypes form distinct phenotypic groups for each of the three glucosinolate ratios (Figure 1, D–F).

One QTL for R3 contains *MAM* while the other QTL for R3 is located near *nga249*, the next fully informative marker below the *MAM* markers. Both markers have an obvious effect on the phenotype (Figure 1, F and G). To test whether *nga249* or the *MAM* markers better explain R3, we used the marker genotypes as predictors of R3 phenotype in separate ANOVAs (Table 1). Genotype information at *nga249* and *MAM* was incomplete, so only those individuals genotyped at both loci ($n = 71$) were used. Both markers had significant effects on R3, but the *MAM* markers effect was stronger, demonstrating that the R3 glucosinolate phenotype is better predicted by *MAM* genotype than by marker *nga249*.

Using the full genetic data, no significant QTL were found for total quantity of aliphatic glucosinolates.

For indolic glucosinolates, both analysis of total indolic glucosinolates and the ratio of total aliphatic to total indolic glucosinolates found one QTL 17.1 cM below the first marker on linkage group 1 (Figure 2, A and B). Given that there was no QTL for total aliphatic glucosinolates and the QTL for the aliphatic to indolic ratio is at the same location as the total indolic glucosinolate QTL, it is likely that the QTL for aliphatic to indolic ratio results from indolic glucosinolate variation. The QTL location does not correspond to the location of any of the four QTL for total indolic glucosinolates that have been found in *A. thaliana* between Ler and Cvi (Kliebenstein *et al.* 2001a). The QTL found here is located near the marker *ICE13* and is centered 3.8 cM above the probable position of *CYP79F1* and *CYP79F2* on the basis of the corresponding position in *A. thaliana*. These genes encode cytochrome P450s that use amino acid substrates in the glucosinolate biosynthetic pathway (Hansen *et al.* 2001; Reintanz *et al.* 2001; Chen *et al.* 2003). *ICE13* is a fully informative marker while the next marker below the cytochrome P450s is not fully informative. Differences in marker informativeness affect Bayesian QTL analysis (Sillanpää and Arjas 1999), and this difference in marker information could have biased the QTL to a position above its real location, supporting the idea that the QTL may be caused by *CYP79F1* or *CYP79F2*.

Cosegregation analysis: Within the Plech crosses each plant was assigned a genotype using the alleles determined by the *MAM* markers. Given two diploid parents, there are a maximum of four alleles that could be monitored. Four crosses contained enough marker variation to determine the segregation of all four alleles, 1 cross identified segregation of three alleles, while the 10 remaining crosses contained only enough marker variation to determine the segregation of two alleles. Analysis by ANOVA found significant effects of *MAM* genotype and FAMILY on all three glucosinolate ratios (Table 2) and six families with distinct associations between genotype and phenotype are displayed in Figure 3. A MANOVA analyzing all three ratios at once was also highly significant at *MAM* (results not shown).

DISCUSSION

Which genes? On the basis of biochemical knowledge of the glucosinolate biosynthetic pathway, we examined ratios of short- to long-chain aliphatic glucosinolates. These ratios are closely related to substrate/product ratios at the step catalyzed by the *MAM*-encoded

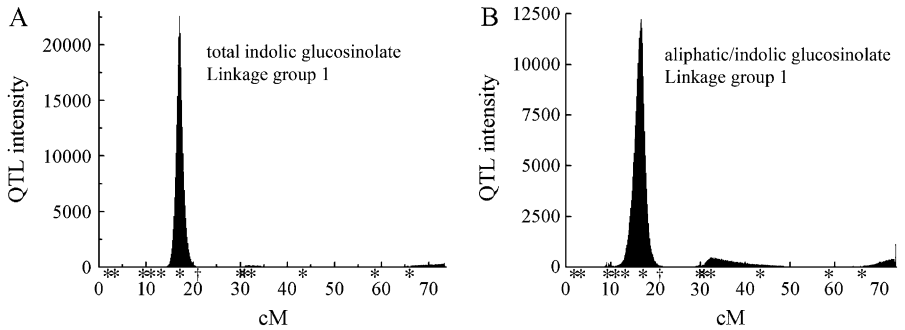

Figure 2.—(A and B) Bayesian posterior QTL intensity graph for total indolic glucosinolates and the ratio of total aliphatic to total indolic glucosinolates. Only linkage groups with QTL are displayed. The position of genetic markers and *CYP79F1* and *CYP79F2* is indicated below the *x*-axis with asterisks and a dagger, respectively.

TABLE 2

Within-population effects of FAMILY and the *MAM* region on glucosinolate ratios R3, R6, and R7

	R7			R6			R3		
Effect tests source	d.f.	SS	*F*	d.f	SS	*F*	d.f.	SS	*F*
FAMILY	14	5.34	4.4****	14	380	9.4****	14	34.2	12.9****
MAM[FAMILY]	24	3.95	1.8*	24	259	3.7****	24	11.9	2.6***

In Plech families, fixed effects ANOVA on R3, R6, and R7 ratios of long/short glucosinolate concentrations shows significant cosegregation of MAM markers with biochemical phenotype. SS, sum of squares. *Significant *F*-values at $P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

enzymes and may be influenced by regulatory or catalytic variation at this locus. We found that alleles in the region containing *MAM* cosegregate with glucosinolate variation both within (Table 2 and Figure 3) and between populations (Figure 1) of *A. lyrata*. Tight linkage of this precise biochemical phenotype to the *MAM*-encoding genes strongly suggests (but does not prove) that *MAM* genes are responsible for this functional variation. Furthermore, no other glucosinolate QTL was detectable with Bayesian QTL analysis (other than a tightly linked QTL detectable only for the R3 ration; see below). If *A. lyrata* has an unknown, unrelated locus that affects these glucosinolate ratios, it is unlikely that it would be located exactly in the *MAM* region. Similarly, the cosegregation study used markers adjacent to the *MAM* locus, which has been shown in *A. thaliana* to affect glucosinolate ratios. Indeed, it has been proposed that one method to identify the gene responsible for a QTL is to find natural genetic variation at a locus that is known to affect the trait in another species (Weigel and Nordborg 2005).

QTL mapping identified a region 20 cM downstream of *MAM* that influences only the R3 glucosinolate ratio (Figure 1). However, more detailed genetic analysis (Table 1) shows that the *MAM* region itself is the best statistical predictor of the R3 glucosinolate phenotype.

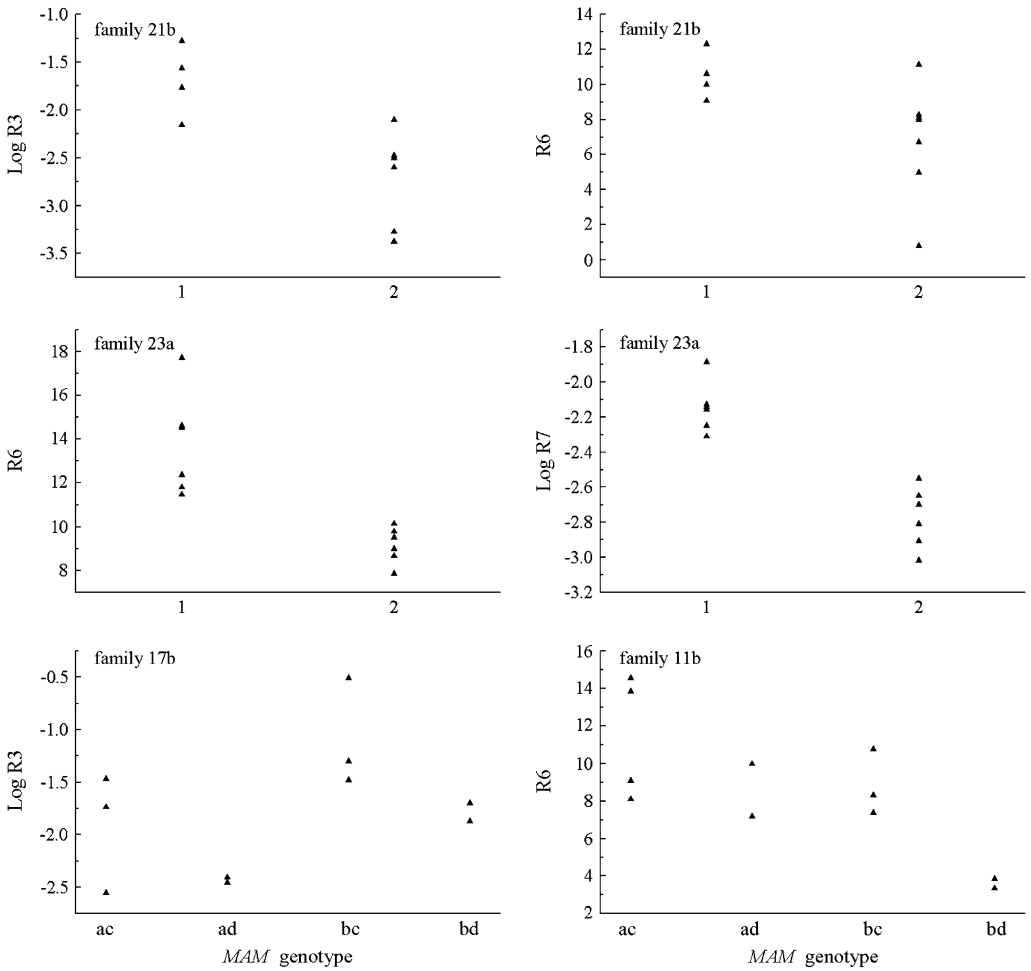

Figure 3.—Effect of *MAM* genotype on glucosinolate ratios for six combinations of glucosinolate ratio and family. Genotype labels are consistent within families, but the relationship between genotypes in different families is unknown.

Still, this linked region deserves further attention, as Kliebenstein *et al.* (2002) found an *A. thaliana* QTL in the homologous region 20 cM from *MAM* near the *EC198L* AFLP marker that influenced resistance to a generalist insect herbivore and aliphatic glucosinolate concentration (but not chain length).

The *MAM* region harbors a QTL for total aliphatic glucosinolates in *A. thaliana* (Kliebenstein *et al.* 2001a; Kroymann *et al.* 2003) and *Brassica napus* (Toroser *et al.* 1995), but not in *A. lyrata*. This is somewhat unexpected since *MAM* allelic variation is present in *A. lyrata* and all aliphatic glucosinolates are processed by *MAM* enzymes. One possible explanation for the lack of a total glucosinolate QTL is that the *MAM* variation sampled in *A. lyrata*, while variable for the end products, may not be variable for overall catalytic rate. Further studies are necessary to infer how *MAM* variation affects total aliphatic glucosinolates *vs.* aliphatic glucosinolate ratios.

Within the *MAM* locus, it is unknown which particular gene family member may be responsible for these biochemical QTL. In *A. thaliana* allelic variation together with multiple *MAM* presence–absence polymorphisms influence variation in chain-elongated aliphatic glucosinolates (Kroymann *et al.* 2003), but further study is required to determine the exact gene or genes responsible in *A. lyrata*.

The total indolic glucosinolate QTL and the aliphatic to indolic glucosinolate ratio QTL may be controlled by *CYP79F1* or *CYP79F2* (Figure 2), but these genes should be mapped in *A. lyrata* to conform colinearity with *A. thaliana*. Furthermore, we cannot rule out an unknown linked locus. Interestingly, these *A. thaliana* cytochrome P450s have *in vitro* substrate specificity for the biosynthesis of aliphatic but not indolic glucosinolates but still have demonstrated effects on indolic glucosinolate concentration *in planta* (Chen *et al.* 2003). Various allelic forms of the enzymes might shunt different amounts of aliphatic glucosinolate intermediates into the biosynthetic pathway, where they might compete with indolic glucosinolate intermediates for continued biosynthesis by less specific enzymes, thereby altering the flux of indolic glucosinolates. Further work is needed to identify the actual causal locus and determine levels of within-population variation.

Genetic basis of quantitative variation: To understand the evolutionary processes responsible for complex trait variation we must elucidate genetic architecture at multiple levels: within populations, among populations, and between species. Standard QTL analysis can identify the genes responsible for phenotypic differences but does not determine the full range of variability within a population or between populations. Conversely, a candidate gene analysis can examine a more diverse sample of variability at a locus but cannot identify unknown loci that may affect the trait of interest. We use an outbred QTL design that allows the monitoring of four segregating alleles in each family, rather than contrasting two alleles as with inbred designs. It has been argued that inbred line crosses are insufficient to determine the genetic architecture of natural populations (Slate 2005) and inbred line crosses can only follow the segregation of two alleles, thus missing much of the natural variation within populations. Thus, outbred designs have an important role for identifying variation in natural populations.

The QTL study of the K × M cross found that the *MAM* region is the only major determinant of variation in aliphatic glucosinolate ratios between these populations. The relatively small sample size used here may have missed QTL of small effects, but these effects, if they exist, are minimal compared to the *MAM* effect. Similarly, the within-population cosegregation study found extensive variation at *MAM* and found that this variation determines aliphatic glucosinolate ratios. There may be other polymorphic loci that affect glucosinolate ratios within this population, but the *MAM* locus clearly has a large effect.

Intraspecific allelic polymorphisms at the *MAM* complex locus control glucosinolate variation in *A. thaliana* (Kroymann *et al.* 2001, 2003), *A. lyrata* (this study), and *B. oleracea* (Li and Quiros 2002). Taken together, the evidence from within a population, between populations, and between species suggests that natural variation in *MAM* is a fundamental determinant of natural variation in aliphatic glucosinolates in multiple species across tens of millions of years.

Variation in aliphatic glucosinolate ratios is heritable within the Plech population, where the plants are attacked by a diverse assemblage of herbivores, including both Brassicaceae specialists and generalists (Clauss *et al.* 2006). Further information on effects and frequencies of *MAM* alleles is required to understand the evolutionary and ecological significance of this within-population variation (Barton and Keightley 2002; Turelli and Barton 2004). However, given the extensive genetic variation found within this population, clearly selection has not fixed *MAM* alleles nor eliminated functionally important genetic variation. Temporally and spatially diverse patterns of natural selection by herbivores may contribute to maintenance of genetic variation at this ecologically important locus.

The authors thank Mikko Sillanpää for extensive support using Multimapper/outbred software. We also thank Helmi Kuittinen for assistance with the genetic data from the K × M cross and two anonymous reviewers for improving the article. We acknowledge the Max Planck Society for financial support.

LITERATURE CITED

Barton, N. H., and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nat. Rev. Genet. **3:** 11–21.

Brown, P. D., J. G. Tokuhisa, M. Reichelt and J. Gershenzon, 2003 Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochemistry **62:** 471–481.

Chen, S., E. Glawischnig, K. Jorgensen, P. Naur, B. Jorgensen *et al.*, 2003 Cyp79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in Arabidopsis. Plant J. **33:** 923–937.

Clauss, M. J., and T. Mitchell-Olds, 2003 Population genetics of tandem trypsin inhibitor genes in *Arabidopsis* species with contrasting ecology and life history. Mol. Ecol. **12:** 1287–1299.

Clauss, M. J., and T. Mitchell-Olds, 2006 Population genetic structure of Arabidopsis lyrata in Europe. Mol. Ecol. (in press).

Clauss, M. J., H. Cobban and T. Mitchell-Olds, 2002 Cross-species microsatellite markers for elucidating differences in population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). Mol. Ecol. **11:** 591–601.

Clauss, M. J., S. Dietel, G. Schubert and T. Mitchell-Olds 2006 Glucosinolate and trichome defenses in a natural Arabidopsis lyrata population. J. Chem. Ecol. (in press)

Dellaporta, S. L., J. Wood and J. B. Hicks, 1985 Maize DNA miniprep, pp. 36–37 in *Molecular Biology of Plants: A Laboratory Course Manual*, edited by R. Malmberg, J. Messing and I. Sussex. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Field, B., G. Cardon, M. Traka, J. Botterman, G. Vancanneyt *et al.*, 2004 Glucosinolate and amino acid biosynthesis in Arabidopsis. Plant Physiol. **135:** 828–839.

Hansen, C. H., U. Wittstock, C. E. Olsen, A. J. Hick, J. A. Pickett *et al.*, 2001 Cytochrome P450 CYP79F1 from Arabidopsis catalyzes the conversion of dihomomethionine and trihomomethionine to the corresponding aldoximes in the biosynthesis of aliphatic glucosinolates. J. Biol. Chem. **14:** 11078–11085.

Karkkainen, K., H. Kuittinen, R. van Treuren, C. Vogl, S. Oikarinen *et al.*, 1999 Genetic basis of inbreeding depression in *Arabis petraea*. Evolution **53:** 1354–1365.

Kliebenstein, D. J., J. Gershenzon and T. Mitchell-Olds, 2001a Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics **159:** 359–370.

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

Kliebenstein, D. J., V. M. Lambrix, M. Reichelt, J. Gershenzon and T. Mitchell-Olds, 2001c Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. Plant Cell **13:** 681–693.

Kliebenstein, D. J., D. Pedersen, B. Barker and T. Mitchell-Olds, 2002 Comparative analysis of quantitative trait loci controlling glucosinolates myrosinase and insect resistance in *Arabidopsis thaliana*. Genetics **161:** 325–332.

Kliebenstein, D. J., J. Kroymann and T. Mitchell-Olds, 2005 The glucosinolate-myrosinase system in an ecological and evolutionary context. Curr. Opin. Plant Biol. **8:** 264–271.

Koch, M., B. Haubold and T. Mitchell-Olds, 2001 Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. Am. J. Bot. **88:** 534–544.

Koornneef, M., C. Alonso-Blanco and D. Vreugdenhil, 2004 Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. **55:** 141–172.

Kroymann, J., S. Textor, J. G. Tokuhisa, K. L. Falk, S. Bartram *et al.*, 2001 A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. Plant Physiol. **127:** 1077–1088.

Kroymann, J., S. Donnerhacke, D. Schnabelrauch and T. Mitchell-Olds, 2003 Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc. Natl. Acad. Sci. USA **100:** 14587–14592.

Kuittinen, H., A. A. de Haan, C. Vogl, S. Oikarinen, J. Leppälä *et al.*, 2004 Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*. Genetics **168:** 1575–1584.

Lambrix, V., M. Reichelt, T. Mitchell-Olds, D. J. Kliebenstein and J. Gershenzon, 2001 The Arabidopsis epithiospecificier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. Plant Cell **13:** 2793–2807.

Li, G., and C. F. Quiros, 2002 Genetic analysis, expression and molecular characterization of *BoGSL-ELONG*, a major gene involved in the aliphatic glucosinolate pathway of Brassica species. Genetics **162:** 1937–1943.

Mikkelsen, M. D., C. H. Hansen, U. Wittstock and B. A. Halkier, 2000 Cytochrome P450 CYP79B2 from *Arabidopsis* catalyzes the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. J. Biol. Chem. **275:** 33712–33717.

Mitchell-Olds, T., 2001 *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. Trends Ecol. Evol. **16:** 693–700.

Reintanz, B., M. Lehnen, M. Reichelt, J. Gershenzon, M. Kowalczyk *et al.*, 2001 A bushy Arabidopsis CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. Plant Cell **13:** 351–367.

Schierup, M. H., 1998 The effect of enzyme heterozygosity on growth in a strictly-outcrossing species, the self-incompatible *Arabis petraea*. Hereditas **128:** 21–31.

Sillanpää, M. J., and E. Arjas, 1999 Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics **151:** 1605–1619.

Slate, J., 2005 Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Mol. Ecol. **14:** 363–379.

Symonds, V. V., A. V. Godoy, T. Alconada, J. F. Botto, T. E. Juenger *et al.*, 2005 Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. Genetics **169:** 1649–1658.

Textor, S., S. Bartram, J. Kroymann, K. L. Falk, A. Hick *et al.*, 2004 Biosynthesis of methionine-derived glucosinolates in Arabidopsis thaliana: Recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain elongation. Planta **218:** 1026–1035.

Toroser, D., C. E. Thormann, T. C. 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.

Turelli, M., and N. H. Barton, 2004 Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and $G \times E$ interactions. Genetics **166:** 1053–1079.

Weigel, D., and M. Nordborg, 2005 Natural variation in Arabidopsis. How do we find the causal genes? Plant Physiol. **138:** 567–568.

Windsor, A. J., M. Reichelt, A. Figuth, A. Svatos, J. Kroymann *et al.*, 2005 Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). Phytochemistry **66:** 1321–1333.

Wittstock, U., and B. A. Halkier, 2002 Glucosinolate research in the *Arabidopsis* era. Trends Plant Sci. **7:** 263–270.

Communicating editor: A. H. Paterson