Supporting Information

Brachi et al. 10.1073/pnas.1421416112

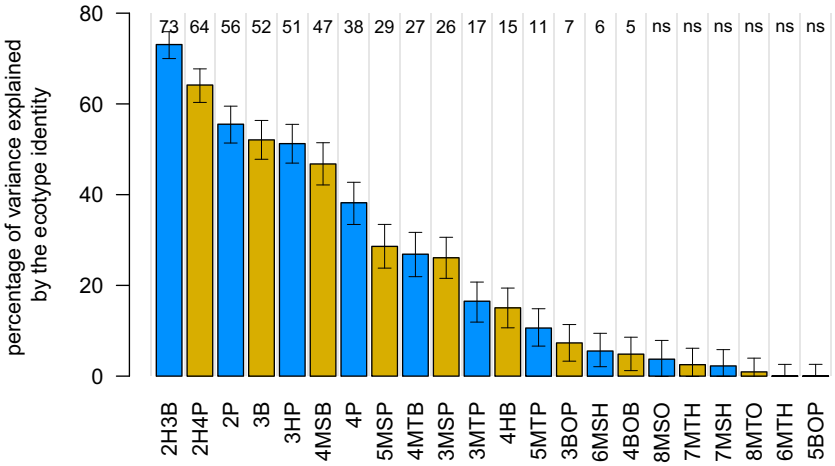

Fig. S1. Broad-sense heritability of the 22 glucosinolate molecules (see Table S1 for abbreviations). The bars represent the proportion of variance explained by the random intercept for the identity of accessions. Variance components were estimated by fitting mixed models (REML), and the error bars represent 95% confidence intervals generated from 1,000 bootstraps.

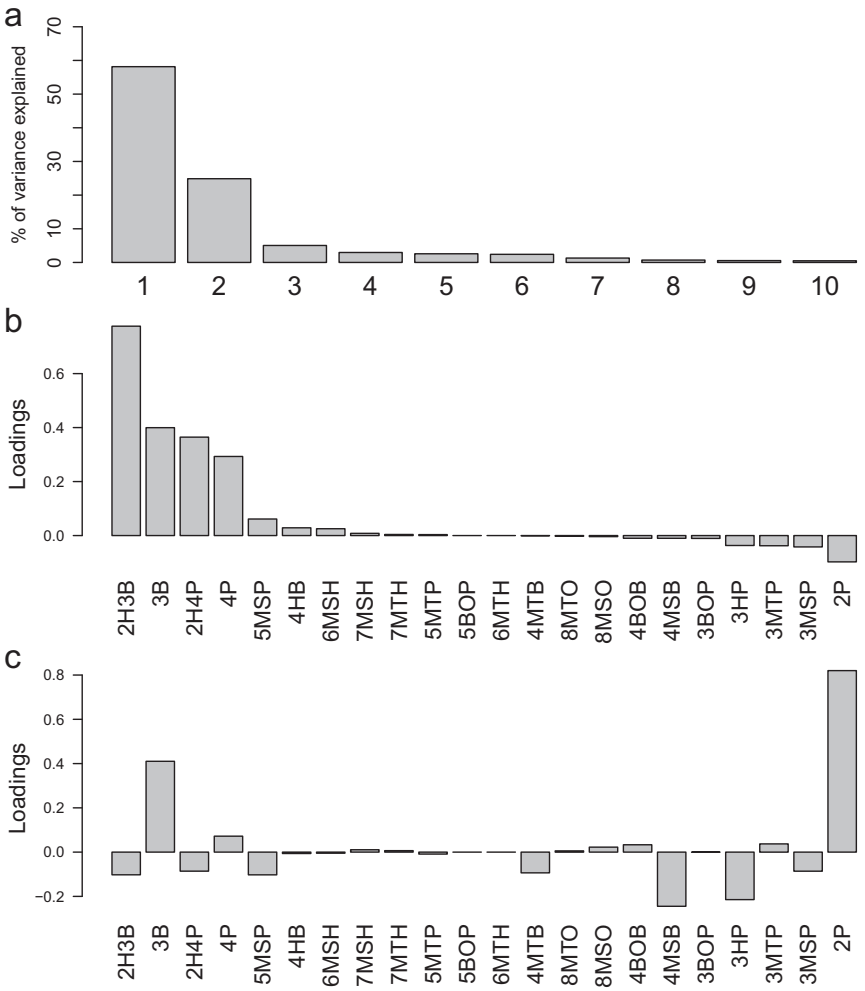

Fig. S2. Principal-component analysis describing variation in the glucosinolate profile in Europe. (*A*) Percentage of variance explained by each of the first 10 components over the total variance they explained. (*B*) Contribution of the glucosinolate molecules to the first principal component. (*C*) Contribution of the glucosinolate molecules to the second principal component.

Fig. S3. Illustration of herbivory categories based on photographs of plants grown in a common garden in Lille (France). 0, no attack visible by the naked eye; 1, percentage of rosette area attacked <10%; 2, percentage of rosette area attacked between 10% and 50%; 3, percentage of rosette area attacked between 50% and 90% (damaged meristem in the rosette center); 4, percentage of rosette area attacked >90% (damaged meristem in the rosette center).

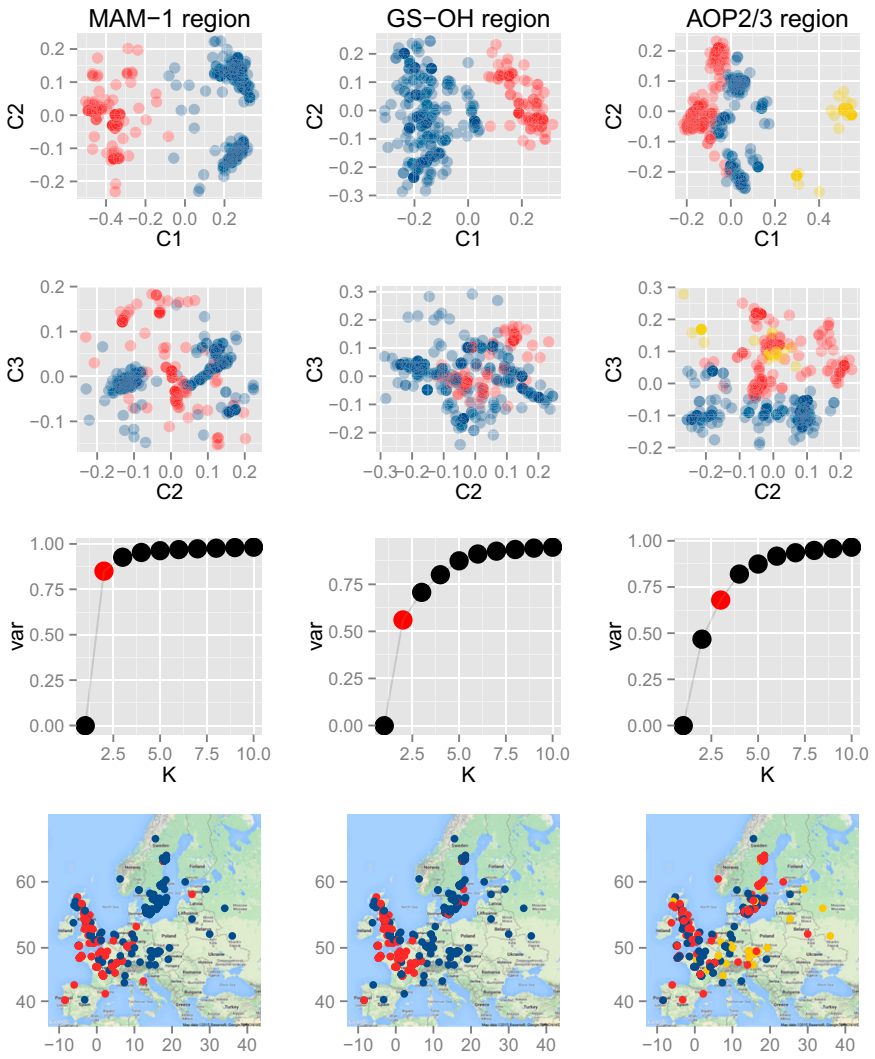

Fig. S4. Allelic variation captured by the SNPs at the loci *MAM*, *GS-OH*, and *AOP* (from *Left* to *Right*). To visualize the allelic variation captured in each region involved in glucosinolate natural variation, we performed multidimensional scaling and *k*-means clustering on the pairwise genetic distance matrix (1 − kinship) between accessions for each genomic region. For each region, we plotted the coordinates of each accession along the first and second components (*Top*) and the second and third components (second from the *Top*) of the multidimensional scaling analysis. In both plots, the colors correspond to genetic clusters. The third graph from the *Top* presents the results of *k*-means clustering for 1–10 clusters, each performed with 1,000 random starts. The *x* axis presents the number of clusters, the *y* axis, the variance explained, and the dots in red mark the number of clusters chosen for representation. The map at the *Bottom* illustrates the geographical distribution of "alleles" defined by the *k*-means clustering.

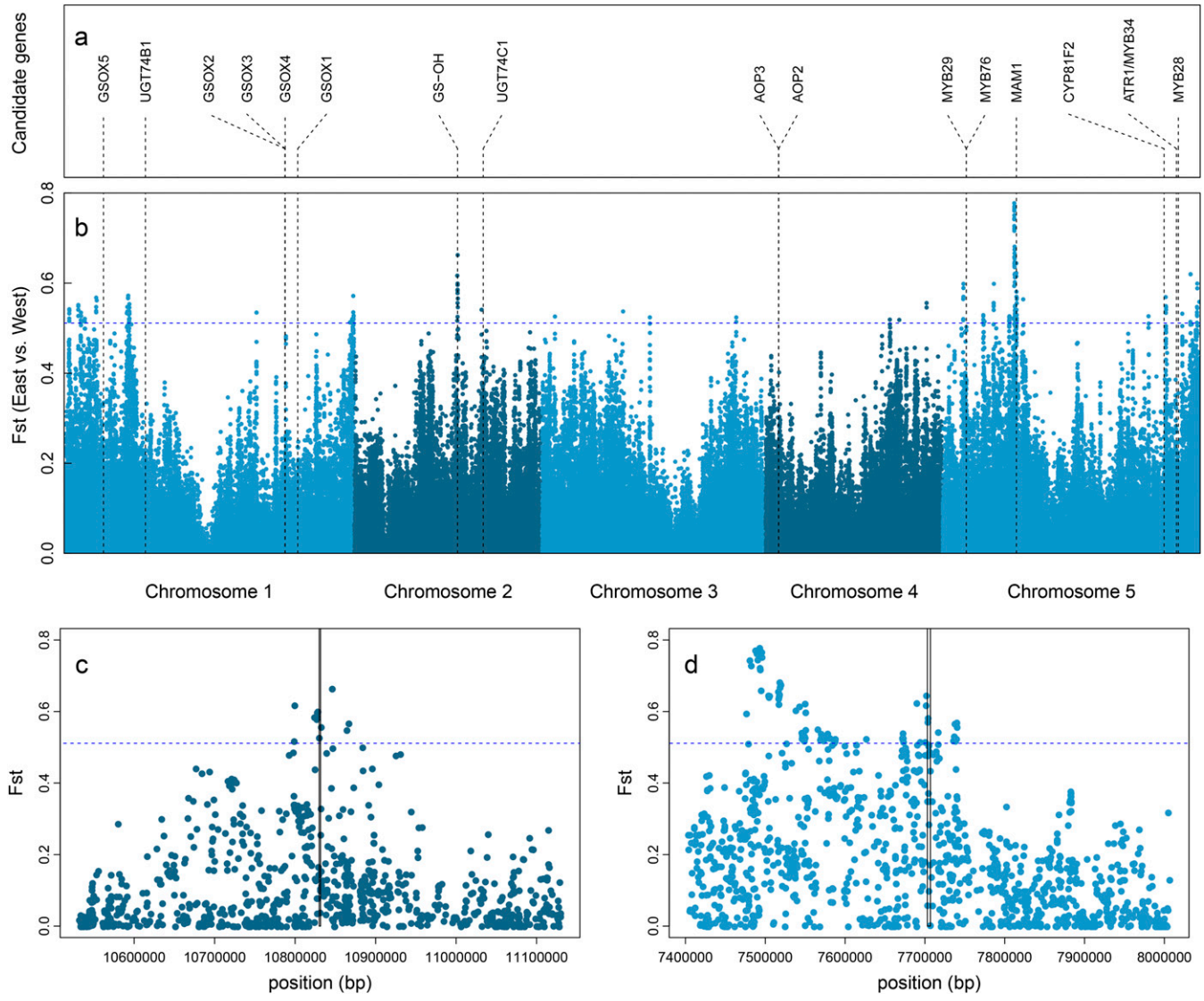

Fig. S5. F_{ST} scan comparing accessions from the east and west of Europe. (*A*) Positions of key glucosinolate biosynthesis genes. (*B*) The populations defined in Horton et al. (1) were merged into two groups, one in the west (British Isles, France, and Iberia) and one in the east (Northwest Europe, South Central, Austria–Hungary, and Fennoscandia). Using this grouping, the F_{ST} scan indicates that the *MAM* and *GS-OH* regions are among the most differentiated genomic regions between Eastern and Western Europe. The blue dashed line represents the 99.9% quantile of the F_{ST} distribution. (*C* and *D*) Zoom on regions 300 kbp on each side of *GS-OH* and *MAM1*, respectively. In both *C* and *D*, the vertical lines represent the limits of *GS-OH* and *MAM1*, respectively.

1. Horton MW, et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44(2):212–216.

Table S1. Twenty-two methionine-derived glucosinolates identified in this study

Name	Short name	*m/z*	*m/z* start	*m/z* stop	RT start	RT stop	IM1	IM2
Allyl (sinigrin)	2P	358.03	357.1	358.8	30	120	13.43	10.666
3-Butenyl	3B	372.04	371.1	372.7	40	150	14.55	10.666
3-Hydroxypropyl	3HB	376.04	375.3	377	20	75	13.49	10.866
4-Pentenyl	4P	386.06	385	386.7	40	200	15.67	10.672
2-Hydroxy-3-butenyl	2H3B	388.2	387.4	388.8	30	120	14.59	10.866
4-Hydroxybutyl	4HB	390.05	389.4	391	30	70	14.61	10.866
2-Hydroxy-4-pentenyl	2H4P	402.05	401.3	403.1	30	100	15.71	10.872
3-Methylthiopropyl	3MTP	406.03	405	406.75	40	120	15.35	15.066
4-Methylthiobutyl	4MTB	420.05	419.3	420.9	70	200	16.48	15.072
3-Methysulfinylpropyl	3MSP	422.02	421.4	422.9	30	60	15.39	15.266
5-Methylthiopentyl	5MTP	434.06	433.3	434.8	150	300	17.6	15.078
4-Methylsulfinylbutyl	4MSB	436.04	435.3	436.7	30	90	16.51	15.272
6-Methylthiohexyl	6MTH	448.08	447.5	448.8	60	280	18.76	15.284
5-Methylsulfinylpentyl	5MSP	450.06	449.5	450.8	40	80	17.63	15.278
7-Methylthioheptyl	7MTH	462.09	461.2	462.8	250	500	19.84	15.09
6-Methylsulfinylhexyl	6MSH	464.07	463.3	464.8	50	120	18.76	15.284
8-Methylthiooctyl	8MTO	476.11	475.1	477	320	500	20.96	15.096
7-Methylsulfinylheptyl	7MSH	478.09	477.5	478.9	60	210	19.88	15.29
3-Benzoyloxypropyl	3BOP	480.06	480	482	240	320	21.27	11.102
8-Methylsulfinyloctyl	8MSO	492.11	491.1	492.85	100	500	21	15.296
4-Benzoyloxybutyl	4BOB	494.07	494.2	496	100	350	22.39	11.108
5-Benzoyloxypentyl	5BOP	508.09	507.4	508.9	180	300	23.51	11.114

m/z is the mass-to-charge ratio of the glucosinolate precursor ions detected in this study. *m/z* start and stop provide the range of masses between which ions were counted. RT start and stop indicate the range of retention time between which the peaks were extracted. IM1 and IM2 provide the height of the predicted isotopic peaks at m + 1 and m + 2 as a percentage of the main peak at *m/z* = m.

Table S2. Geographical variation of the glucosinolate profile

Parameter	Estimate	SE	*t* value	*Pr*(>*t*)
Intercept	6.052e+00	7.415e+00	0.816	0.414752
Longitude	6.071e-01	1.587e-01	3.827	0.000144***
Latitude	−3.979e-02	2.966e-01	−0.134	0.893335
Longitude^2	−3.178e-02	4.547e-03	−6.990	7.59e-12***
Latitude^2	−1.026e-03	2.957e-03	−0.347	0.728664
Longitude ∗ latitude	−1.482e-02	3.016e-03	−4.913	1.17e-06***
Longitude^2 ∗ latitude^2	1.199e-05	1.585e-06	7.566	1.52e-13***

Results of the multiple regression investigating the relationship between the first component of the glucosinolate profile and the location of origin of the accessions. Latitude^2 and Longitude^2 are second-order terms included in the model and accounting for nonlinear relationships between glucosinolate profile and spatial coordinates. Interactions are marked by ∗. Significance: ***$P < 0.001$. The adjusted r^2 for the model was 0.6.

Table S3. Genetic variation of an estimate of lifetime fitness: Random effects

Groups	Name	Variance	SD	Correlation	
Random effects					
Accession	Treatment – August 2010 (intercept)	2.122e+02	1.457e+01		
	Treatment – September 2010	7.282e+01	8.533e+00	−0.99	
	Treatment – September 2011	3.426e+02	1.851e+01	−0.51	0.42
Accession	(Intercept)	3.426e-12	1.851e-06		
Block	(Intercept)	1.426e+00	1.194e+00		
ε		5.021e+02	2.241e+01		

Tables S3 and S4 present the fit of the mixed model following Eq. **5**.

Table S4. Genetic variation of an estimate of lifetime fitness: Fixed effects and correlation of fixed effects

Groups	Estimate	SE	*t* value
Fixed effects			
Treatment – August 2010 (Intercept)	47.867	1.405	34.06
Treatment – September 2010	−3.737	1.727	−2.16
Treatment – September 2011	−10.533	1.925	−5.47
Correlation of fixed effects			
	Treatment – August 2010 (intercept)	Treatment – September 2010	
Treatment – September 2010	−0.721		
Treatment – September 2011	−0.660	0.483	

Tables S3 and S4 present the fit of the mixed model following Eq. **5**.

Table S5. Negative effect of the level of herbivore damage on plant fitness

Parameter	Estimate	SE	*t* value	*P* value	Significance
Intercept	3.61	3.30e-02	109.422	<2.00e-16	***
Distance	−3.52e-04	4.17e-05	−8.446	2.38e-15	***
Herbivory	−1.15e-01	4.82e-02	−2.394	0.0174	*

Herbivore damage had a significant effect on fitness in a multiple regression including the geographical distance between the experiment and the original site of collection of each accession ("distance"). Significance: ***$P < 0.001$; *$P < 0.05$. The adjusted r^2 for the model was 0.24.

Other Supporting Information Files

Dataset S1 (XLSX)