Table S1

Origins of plant samples

| Individual/ Abbreviation | ABRC[a] stock number | $n_{genic}$[b] | $n_{promoter}$[c] | Origin | Collector |
|---|---|---|---|---|---|
| *Arabidopsis lyrata lyrata* | | | | | |
| NYA | | 2 | 2 | New York, USA[h] | |
| PI11.5 | | 1 | 2 | Presque Isle State Park, PA, USA | Janette Steets |
| RS3.3, RS9.1 | | 4 | 3 | Racoon Creek State Park, PA, USA | Janette Steets |
| GM4.5 | | 2 | 1 | Grand Mere State Park, MI, USA | Janette Steets |
| WIA, WIC | | 4 | 2 | Wisconsin, USA[h] | |
| *A. lyrata petraea* | | | | | |
| EsA, EsB | | 4 | 4[d] | Esja Mountain, Iceland[i] | |
| ReB | | 2 | 1 | Reykjavik, Iceland | M. Schierup |
| BrA, BrB | | 2 | 2 | Braemar, Scotland | Richard Ennos |
| ExA, ExZ | | 4[e] | 2 | Exeter, England | Mark MacNair |
| PlA, PlB | | 2 | 2 | Plech, Germany[h] | T. Mitchell-Olds |
| Ka | | 2 | 2 | Karhumäki, Russia | O. Savolainen |

*A. thaliana*

| | | | | | |
|---|---|---|---|---|---|
| Yo-0 | CS1622 | 1 | 1 | Yosemite National Park, California, USA | |
| Tol-6 | CS8026 | 1 | 0 | Toledo, Ohio, USA | |
| Durham | | 1 | 1 | Durham, North Carolina, USA | D. Wolf, N. Takebayashi |
| Mv-0 | CS1387 | 1 | 0 | Martha's Vineyard, Massachusetts, USA | |
| Co-2 | CS1086 | 1 | 1 | Coimbra, Portugal | |
| Pla-2 | CS6916 | 1 | 1 | Playa de Aro, Spain | |
| Se-0 | CS1502 | 0 | 1 | San Eleno, Spain | |
| Bur-0 | CS1028 | 0 | 1 | Burren, Ireland | |
| Mc-0 | CS1362 | 0 | 1 | Mickles Fell, UK | |
| Bu-2 | CS1008 | 1 | 0 | Burghaun, Rhön, Germany | |
| Col[f] | | 1 | 1 | Ecotype Columbia, Germany | |
| Bs-5 | CS1000 | 1 | 1 | Basel, Switzerland | |
| Zdr-6 | CS22589 | 0 | 1 | Žďárec, Brno, Czech Republic | |
| Ct-1 | CS1094 | 1 | 1 | Catania, Italy | |
| Mir-0 | CS1379 | 0 | 1 | Miramare, Trieste, | |

| | | | | | |
|---|---|---|---|---|---|
| | | | | Italy | |
| Hau-0 | CS1220 | 0 | 1 | Hauniensis, Denmark | |
| Oy-0 | CS1436 | 1 | 0 | Oystese, Norway | |
| Ost-0 | CS1430 | 1 | 1 | Osthammar, Sweden | |
| Ws-0 | CS22623 | 0 | 1 | Wassilewskija, Russia | |
| N14 | CS22492 | 1 | 0 | Sampo Mountain, Russia | |
| Sha | CS57924 | 0 | 1 | Pamiro-Alay, Tadjikistan | |
| Kas-2 | CS6751 | 0 | 1 | Kashmir, India | |
| Tsu-0 | CS1564 | 0 | 1 | Tsu, Japan | |
| *A. halleri* | | | | | |
| | | 1 | 0 | Morgante du Nord, France | P. Saumitou-Laprade |
| *Boechera stricta* (= *Arabis drummondii*) | | | | | |
| | | 1 | 0 | Galena, Alaska, USA | Matt Olson |
| *Brassica rapa* ssp. *Pekinensis*[g] | | | | | |
| | | 1 | 0 | | |

[a] Arabidopsis Biological Resource Center.

[b] Number of alleles sampled for the analysis of *MEA* genic region.

[c] Number of alleles sampled for the analysis of *MEA* promoter region.

[d] One of the alleles was excluded in the analysis of the 5' flanking gene due to a low sequence quality at the 3'-end.

[e] One of the alleles comes from cDNA, therefore, this allele was used only for the OmegaMap analysis.

[f] Downloaded from Genbank (accession number AC022521).

[g] Downloaded from Genbank (accession number AC189527).

[h] Seeds were kindly donated by Tom Mitchell-Olds.

[i] Seeds were kindly donated by Stephen Wright and Deborah Charlesworth.

Table S2

Estimates of population recombination parameters per base pair, ρ (x $10^{-3}$). We used LDhat 2.1 (McVean, Awadalla, and Fearnhead 2002) to estimate population recombination parameters, $\rho = 4 N_e r$, where $r$ is the rate of recombination per neighboring bases. Since LDhat estimates the population recombination parameters for the whole region, we divided the output by aligned sequence length minus 1 to convert it to be $\rho$ per neighboring bases.

| | 5' gene + *MEA* promoter | *MEA* genic region |
|---|---|---|
| *A. lyrata* | 0.00 | 2.98 |
| *A. l. lyrata* | 0.00 | 6.75 |
| *A. l. petraea* | 0.00 | 3.68 |
| *A. thaliana* | 3.28 | 1.19 |

**Supplement A: Details of OmegaMap analysis**

We used 1,000,000 Markov-chain Monte Carlo samplings, and the first 50,000 iterations were discarded as a burn-in. Two chains were run from different initial parameter values and the convergence of the two chains were confirmed. Two chains were thinned to one sample every 1000 iterations and merged to obtain the posterior distribution. For *A. lyrata*, we assumed that all codons have equal equilibrium frequencies. For *A. thaliana*, we used the empirically estimated codon frequencies from the entire genome as the equilibrium frequencies. The parameters estimated are the selection parameter ($\omega = d_n/d_s$; ratio of nonsynonymous to synonymous substitutions), the rate of synonynmous transversion ($\mu$), the transition transversion rate ratio ($\kappa$), the rate of insertion/deletion ($\phi$), and population recombination rates ($\rho$). The recombination rates and $\omega$ were allowed to vary in different regions along the sequence. The method assumes that the neighboring codons have similar properties, and that heterogeneity in $\omega$ and $\rho$ show block-like structure. The lengths of each block is randomly determined for each iteration, and we set the parameters so that the average block size is 10 codons for $\omega$, and 100 codons for $\rho$. Larger average block sizes reduce the computational time, however, positively selected sites will not be detected if only a few sites are actually under the selection. We set a larger block for recombination, but we set a smaller block so that we do not miss such positively selected regions of small sizes.

Because OmegaMap is a Baysian analysis, the prior probability distributions have the potential to influence the posterior probability for each parameter estimate. We used three sets of prior probability distributions to assess the sensitivity of our results to these priors (table S3). For $\omega$, we used inverse, exponential or gamma distributions. An

inverse distribution is a uniform distribution on the logarithmic scale, which represents the assumption that we do not have any strong prior information other than setting the lower and upper bound to the possible parameter. An exponential distribution represents the belief that the majority of codons are conserved and under negative selection. The gamma distribution that we used has a similar shape to the exponential, but it represents the belief that codons with intermediate values of ω ($0 < \omega < 1$) may be more common than invariable codons ($\omega = 0$). Note that we have set the prior distribution of ω, so that the mean is $\omega = 1$ (neutrality) for prior set A and C and $\omega < 1$ for set B. For the other parameters (μ, ϕ, ρ, and κ), we used inverse, improper inverse, exponential, or exponential ratio distributions (table S3), following (Huelsenbeck and Dyer 2004). An "improper" inverse function, is similar to an inverse distribution, except that any positive value is possible.

Table S3

Prior distibutions used for OmegaMap analysis

| **Parameter** | **Prior set A** | **Prior Set B** | **Prior Set C** |
|---|---|---|---|
| μ | Improper inverse | Exponential (mean 0.1) | Exponential (mean 0.01) |
| κ | Improper inverse | Exponential Ratio (mean undefined; median 1) | Exponential Ratio (mean undefined; median 1) |
| ϕ | Improper inverse | Exponential (mean 0.01) | Exponential (mean 0.1) |
| ρ | Inverse (Range $10^{-6}$ – 50; mean 2.82) | Exponential (mean 1) | Exponential (mean 0.01) |
| ω | Inverse (Range 2 x $10^{-9}$ – 23; mean 0.99) | Exponential (mean 0.5) | Gamma (shape 1.5, scale 0.66; mean 0.99) |

Huelsenbeck, J. P., and K. A. Dyer. 2004. Bayesian estimation of positively selected sites. J. Mol. Evol. **58:** 661-672.

McVean, G. A. T., P. Awadalla and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:**1231-1241.