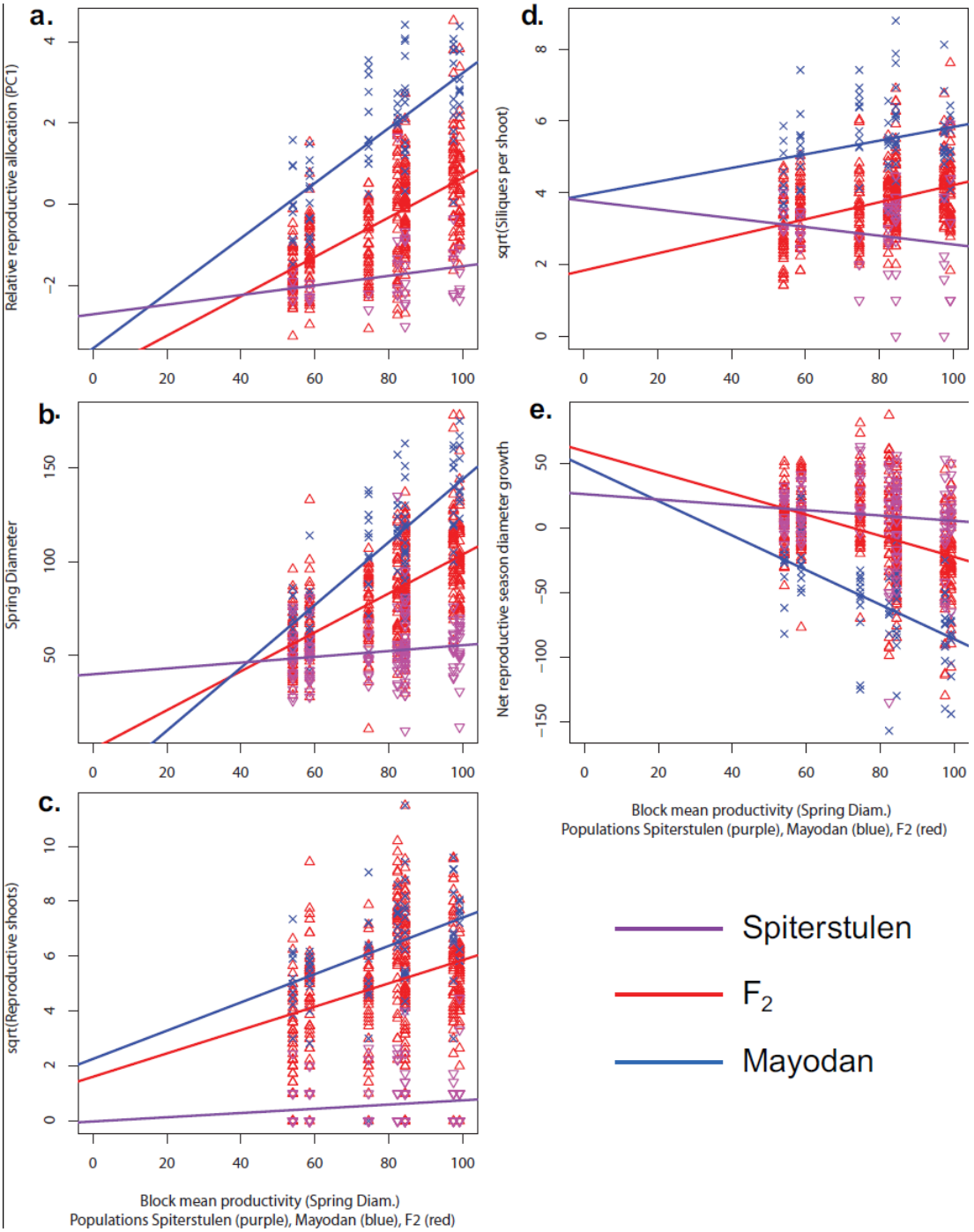

Figure S1 Phenotypic plasticity of populations to variation in block productivity. Plasticity of vegetative and reproductive traits to variation in block productivity, measured as mean block spring diameter, in North Carolina for Spiterstulen, Mayodan, and F2 populations. (**A**) PC1 ($P < 0.001$ for population effects on plasticity); (**B**) spring diameter ($P < 0.001$); (**C**) reproductive shoots (with square root transformation, $P = 0.0077$); (**D**) siliques per shoot (with square root transformation, $P = 0.0020$); and (e) net reproductive season diameter growth ($P < 0.001$). Main effects of block productivity were highly significant for all traits ($P < 0.001$). Estimates are from a mixed model, with population and block productivity treated as fixed effects, and family (or F2 reciprocal) as random effects.

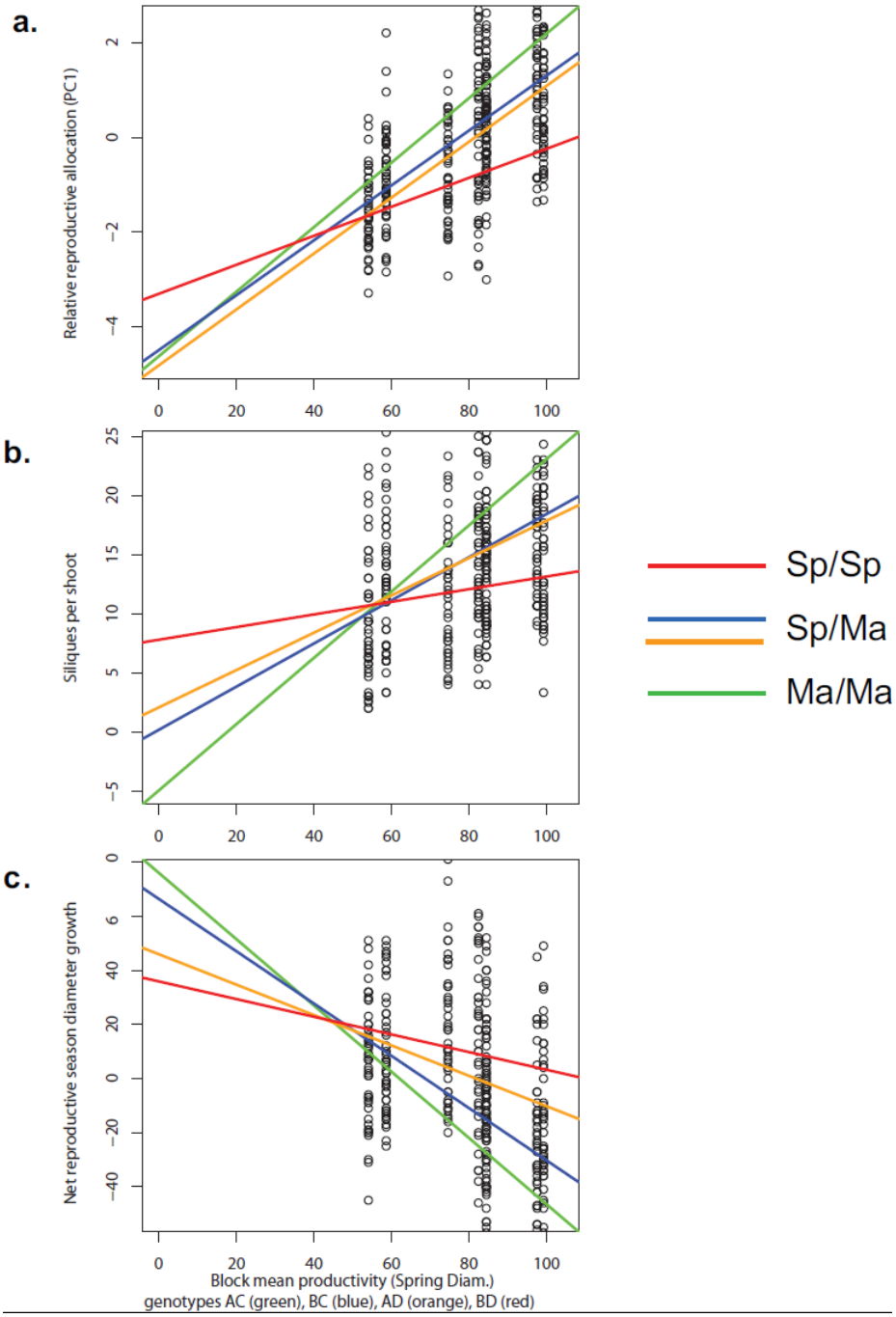

Figure S2 Phenotypic plasticity of LG2 QTL region to variation in block productivity. Plasticity of vegetative and reproductive traits to variation in block productivity, measured as mean block spring diameter, in North Carolina for genotypic classes in the LG2 QTL region: (**A**) PC1 ($P < 0.017$ for QTL effects on plasticity); (**B**) siliques per shoot ($P = 0.035$); and (**C**) net reproductive season diameter growth ($P = 0.038$). QTL genotype probabilities for each F_2 individual were estimated in R/qtl, and coefficients for Spiterstulen homozygotes (red), Mayodan homozygotes (green), and the two heterozygous genotype classes (blue and orange) were estimated from QTL models with block productivity included as an interactive covariate with QTL genotype, using fitqtl with Haley-Knott regression. QTL genotype effects on the slope of the response were significant for each trait ($P < 0.05$).

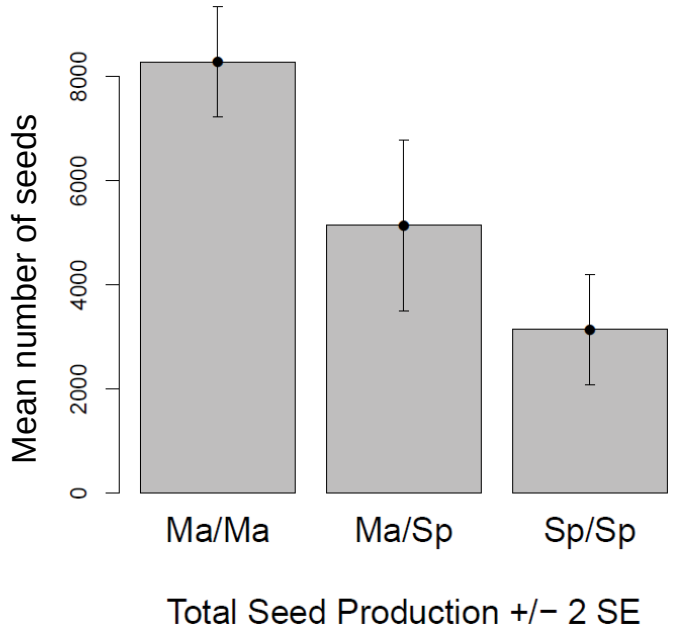

Figure S3 Reproductive output effects of LG2 QTL region in North Carolina. Mean total reproductive output for Spiterstulen homozygotes, heterozygotes, and Mayodan homozygotes at the marker closest to the LG2 QTL peak at the North Carolina field site. Reproductive output was calculated as reproductive shoots x siliques per shoot x seeds per silique.

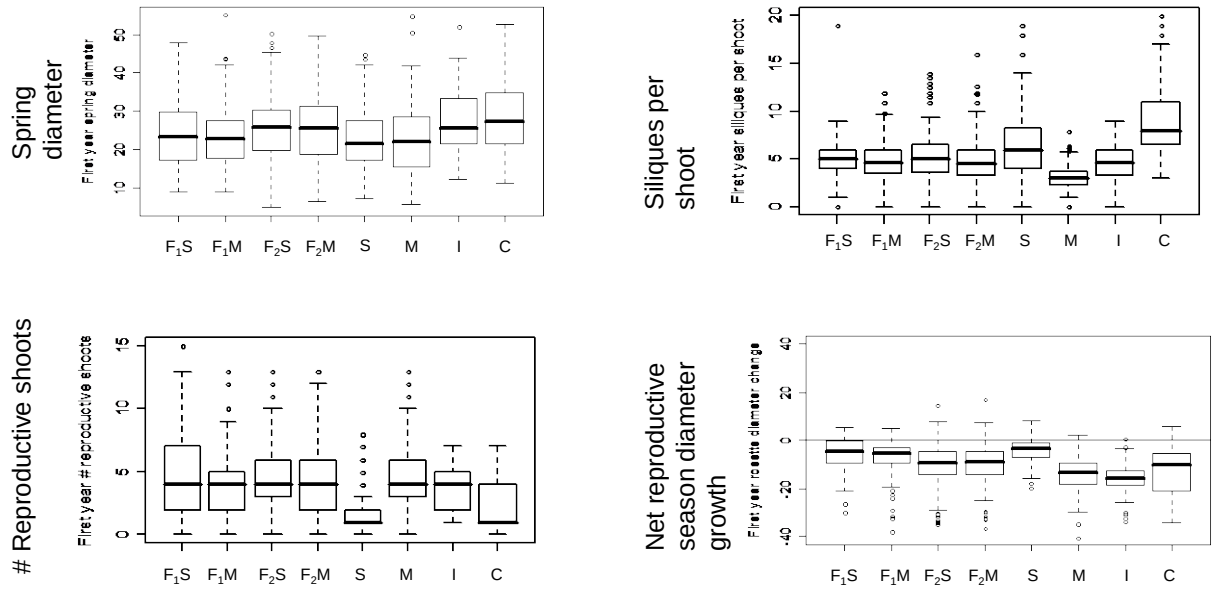

Figure S4 Distributions for vegetative and reproductive traits in Norway. Boxplots show the phenotypic distributions for spring diameter (mm), number of reproductive shoots, siliques per shoot, and net reproductive season diameter growth (mm) for the Norway field site. Groups are F_1 plants (F_1S and F_1M) and F_2 plants (F_2S and F_2M) with Spiterstulen and Mayodan cytoplasm, respectively; and population samples from Spiterstulen (S), Mayodan (M), Ithaca (I), and Chena River, Alaska USA (C).

File S1

Supplemental Methods

Phenotypic plasticity

Differences in phenotypic plasticity between the Mayodan, Spiterstulen and F_2 populations were tested using mixed-effects models in the lmer function from the lme4 package in R. We used block mean spring rosette diameter as measure of block productivity, and tested models in which the productivity covariate was either nested within population (separate slopes and intercepts model) or estimated as a non-nested effect separate from population (single slope, separate intercepts model). A significant likelihood-ratio test for these two models indicated significant differences in the slope of the productivity regression and hence significant differences among populations in plasticity. A likelihood-ratio test comparing the single slope, separate intercepts model with a model lacking the productivity covariate was used to test for overall phenotypic plasticity across populations. For these analyses, full-sib family (or reciprocal in the case of the F_2 population) was included as a random effect. Values of reproductive shoots and siliques per shoot were square-root transformed to improve the residual normality and homoscedasticity of the data.

Outcross F_2 analysis of QTL data

Due to sequence variation both within and between populations in our outcross design, markers segregated in multiple configurations, including fully-informative, F_2 configuration at the population level, and backcross configuration (i.e. informative for only one F_1 parent). R/qtl treats marker-trait data from outcross designs as 4-way data, with four possible genotypic classes at each locus (M_1M_2, S_1M_2, M_1S_2, and S_1S_2), representing Mayodan (M) and Spiterstulen (S) alleles from the two different F_1 parents (M_1S_1 and M_2S_2). Deviations from the mean of the local homozygous class were used to estimate the genotypic effects of the other three classes. To estimate the effects of detected QTL on individual traits, we used the calc.genoprob function to generate pseudomarkers at 2-cM intervals throughout the genome. In the 4-way cross setting, pseudomarker scores consist of the genotype probabilities for the M_1M_2, S_1M_2, M_1S_2, and S_1S_2 classes, on which the phenotypic effects are regressed.

To analyze these data as an outcross F_2 design involving parental crosses between two populations, we reparameterized the R/qtl output to estimate a, d, and i coefficients, which estimate additive effects of Mayodan vs. Spiterstulen homozygotes, dominance effects of Mayodan/Spiterstulen heterozygotes, and deviations of the two heterozygous classes from their mean, respectively. In order to analyze phenotypic effects of QTL in terms of the Mayodan vs. Spiterstulen

allelic origins, we reparameterized pseudomarkers at QTL peak locations by partitioning them orthogonally into additive (a = [$p(M_1M_2) - p(S_1S_2)$]), dominance (d = ½[$p(S_1M_2) + p(M_1S_2) - p(M_1M_2) - p(S_1S_2)$], and heterogeneity ($i$ = [$p(S_1M_2) - p(M_1S_2)$]) terms. We then used the lm function in R (or the glm function for binary traits) to estimate the a, d, and i coefficients. We also tested for block effects and phenotypic plasticity of QTL effects in an analysis similar to the one used with the population data as described above. We compared models in which block mean productivity was included as an interactive covariate with the QTL genotype (separate slopes and intercepts for each genotype), was included as an additive covariate only (single slope with separate intercepts), or was excluded from the model. We developed customized R functions to conduct these analyses.

Since the correlation structure among the vegetative and reproductive traits was suggestive of resource allocation trade-offs, and thus structured pleiotropy for genes associated with these traits, we considered pleiotropy to be highly probable at the QTL level. Thus, if a QTL location had genome-wide significance for any one trait or for PC1, we estimated its effects on each of the other traits. We considered these effects to be significant if an interval of twice the standard error around the estimate of a did not include zero and the deviation was in the expected direction, since these are tests of single loci and no longer genome-wide tests.

Structural equation model optimization

To search for optimal structural equation models in each of the two field environments, we started with a null model containing only QTL genotype-to-trait paths. We then added causal paths from traits expressed earlier in development to later traits, one path at a time. We retained the path that improved the model fit the most (based on Akaike's Information Criterion, AIC) at each step until no further improvement was obtained or until there was only one residual degree of freedom in the model. We then removed all QTL-to-trait paths for which the P-values of the a, d, and i terms were all greater than 0.05, and retested the model to verify that the AIC value was lower and/or the p-value for model fit was higher. To control for possible confounding effects of environmental covariation, we also tested the significance of paths from block mean productivity to traits, and retained productivity-to-trait paths that were significant. A model P-value greater than 0.05 was interpreted as evidence that the model adequately explained the covariance structure of the marker-trait data. We also examined the model goodness-of-fit and adjusted goodness-of-fit indices (GFI and AGFI), which are analogous to R^2 and adjusted R^2 values, respectively, in linear models. The signs and sizes of standardized path coefficients were used to compare estimated path effects with those predicted under the three cause-effect models described in Figure 1.

SEM accommodates the use of latent variables, estimated as factors of the correlation structure of observed variables, to represent unobserved variables that may underlie relationships among observed traits. To evaluate whether the

data are consistent with the hypothesis that net winter diameter growth acts as a surrogate for vegetative rosette branching (model shown in Figure 1c), we also tested a model with a latent variable ("branching", placed in quotes to indicate the trait is unmeasured) upstream of all measured traits except fall diameter. We tested models with paths from all QTL to "branching" and to all of the measured traits except net reproductive season diameter growth; from "branching" to the measured traits (in place of paths from net winter diameter growth); and additional trait-to-trait paths corresponding closely to traits from the best-fitting models without the latent variable. Degrees-of-freedom and structural limitations constrained the number and locations of paths that could be estimated simultaneously. Paths were added and deleted one at a time to obtain the best-fitting estimable model. Finally, non-significant QTL effects were removed, and additional paths from QTL to net reproductive season diameter growth were tested for significance.

We treated all exogenous variables (those strictly upstream of other traits), including block mean productivity, in the SEM as fixed effects without an error variance. Treating exogenous variables as random effects instead would add residual degrees of freedom to the model. With random-effects QTL genotypes, however, covariances between QTL genotypes (including the *a*, *d*, and *i* coefficients for individual QTL) would contribute to model fit, but these covariances are not of interest in the models and have expected values of zero for unlinked QTL lacking segregation distortion. This in turn would reduce the contribution of the relevant parameters (QTL-trait and trait-trait path coefficients) to the overall model fit and inflate the estimated fit of tested models. In practice, we have found that parameter estimates and relative fit of different models are very similar under both approaches. Consequently, we have preferred to use the more conservative estimates of model fit obtained with fixed exogenous variables.

File S2

Supplemental Results

Rosette growth rates and growth patterns both affect spring diameter

Our alternative models to explain coordinated variation in vegetative and reproductive traits (Figure 1a-c) include both resource acquisition and rosette branching as potential explanations for variation in pre-reproductive rosette diameter. Therefore, we investigated the extent to which variation in spring rosette diameter was a function of biomass accumulation (resource acquisition) vs. differences in rosette architecture. Spring diameter at the North Carolina site was positively correlated with reproductive shoots and siliques per shoot but negatively correlated with net reproductive season diameter growth, both in the population means and in the population and F_2 PC1 patterns. This result is inconsistent with the hypothesis that spring diameter is a measure of resource acquisition, in which case we predicted it would be positively correlated both with subsequent reproductive and vegetative growth (Figure 1a; Houle 1991; van Noordwijk and de Jong 1986). These results are instead consistent with the model in which pre-flowering rosette diameter growth is affected by rosette branching (Figure 1c).

We also found that the timing of pre-reproductive rosette growth was heterogeneous between populations, which also suggested that spring diameter has a developmental component. Spiterstulen plants showed a net reduction in rosette diameter over the winter while the other populations showed varying degrees of diameter increase (Figure 4). We observed that rosettes on Spiterstulen plants in particular had become extensively branched and compact over the winter (as seen for example in Figure 3, upper left). We have observed over the course of these studies that branched rosettes tend to produce shorter leaves, suggesting that increased overwinter vegetative branching might explain the overwinter diameter reductions.

However, the eight replicated blocks in the North Carolina field site differed substantially in mean spring diameter. The block mean values for reproductive shoots and siliques per shoot increased significantly with mean spring diameter ($P < 0.001$ for both traits). Thus, variation in pre-flowering rosette diameter also appears in part to reflect differences in resource acquisition associated with site productivity in each block.

Mayodan, Spiterstulen, and F_2 populations also differed significantly from each other in their responses to differences in block productivity ($P < 0.01$ for population x block productivity interactions for all traits). In Mayodan plants, vegetative and reproductive traits as well as PC1 were highly responsive to variation in block productivity, but Spiterstulen plants showed almost no response to block productivity, and F_2 plants were intermediate (Figure S1). This pattern of variation in response to block productivity was also reflected at the QTL level. Mayodan alleles in the LG2 QTL region were significantly more

responsive to variation in block mean productivity than Spiterstulen alleles for PC1 and two of the individual traits, consistent with the population differences (Figure S2).

File S3

Summary of structural equation model runs for North Carolina field data (first worksheet) and Norway field data (second worksheet)

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.151803/-/DC1

All marker data, trait data, and scripts used for data analysis have been deposited in the Dryad Repository: http://dx.doi.org/10.5061/dryad.1k4gq.