Supporting Information

Cook et al. 10.1073/pnas.1118364109

SI Text

Analysis: Chinnor. Based on our P value selection method, we identified 275 spring-only responders and 70 (18.2% of total species) divergent responders at Chinnor. Using the Akaike Information Criterion corrected for small sample sizes (AICc) method, we identified 298 spring-only responders and 47 (12.2%) divergent responders. For our quantitative analysis, we used three statistical tests: a two-sample Student's t test for differences in the mean (accounting for unequal variances), the Kolmogorov–Smirnov (KS) goodness-of-fit test, and the Wilcoxon rank sum (WR) test for differences in the median. Both the KS and WR tests are nonparametric tests, and these are necessary to include because of the relative nonnormality of the underlying data distributions. These nonparametric tests, however, have less power to detect significant differences, and many parametric tests are robust to deviations from nonnormality given sufficient sample sizes (1). Thus, to complement this analysis, we also include results from the traditional Student's t test.

For the 70 P value-selected divergent responders, all three tests showed significant differences in the FFD trends between the spring-only species and the divergent species (Table S1): Mean FFD trends are $-0.14/-0.07$ d·y^{-1} and median trends are $-0.12/-0.04$ d·y^{-1} for the spring-only/divergent species, respectively. The KS and WR tests also show highly significant differences for the 47 AICc-selected species, and results are nearly identical to those selected using the P value selection criteria: mean FFD trends based on the AICc selection are $-0.14/-0.08$ d·y^{-1} and median FFD trends are $-0.11/-0.04$ d·y^{-1} for the spring-only and divergent species, respectively. Taken together, the results strongly support the idea that species with significant vernalization sensitivities are not advancing their flowering over time as rapidly as the spring-only responders. Differences in the magnitude of the spring warming sensitivities between the spring-only and divergent responders are not significant ($P \leq 0.10$) in any of the three tests (Table S2), indicating that both species groups are equally sensitive to spring warming and can be classified as climate responders, despite significant differences in the magnitudes of flowering trends over time.

For our spring-only responders, we use our spring warming sensitivity models to hindcast trends in FFD and compare against observations (Fig. S5). In general, our model predicted that FFD trends match well with the observations, although our model misses the minority of spring-only species that have delayed their flowering. Our model generally reproduces the major features of the observed trends distribution, including the negative central tendency, indicating a general advance in FFD over time in the spring-only responders. The mean and median modeled trends for the 275 P value-selected species are $-0.16/-0.14$ d·y^{-1}; for the 298 AICc-selected species, the mean and median trends are $-0.16/-0.14$ d·y^{-1}. Our modeled distribution has lower variance and underpredicts the proportion of species delaying their flowering. This is not surprising, given that our model is driven only by a single parameter (warming) and, by its nature, will underpredict the observed variance. Differences between observed and predicted trends for these species are insignificant based on the Student's t test but are significantly different at the $P \leq 0.05$ level using the WR test (Table S3). This is likely attributable to the inability of our model to predict the delaying trends. Our large sample size ($n > 200$) and small variance result in the sample mean being less sensitive to the delaying species, leading to an insignificant difference based on our Student's t test. Median

trends, however, are affected by these species, and this results in the significant difference in medians based on the WR test.

We repeat the observed vs. model trend comparison for the divergent responders, comparing observed trends against modeled trends using spring warming sensitivities only and spring warming + vernalization sensitivities (Table S4). Using spring warming sensitivities only, we predict mean/median FFD trends of $-0.16/-0.13$ d·y^{-1} for the 70 P value-selected divergent species and $-0.17/-0.15$ d·y^{-1} for the 47 AICc-selected species. These trends are significantly different from the observations and, in fact, are quite close to the trends observed for the spring-only responders. When the models are refit to incorporate the additional vernalization sensitivities, modeled trends are much weaker and agree much better with (i.e., are not significantly different from) observations: mean/median FFD trends are $-0.05/-0.03$ d·y^{-1} for the 70 P value-selected species and $-0.05/-0.03$ d·y^{-1} for the 47 AICc-selected species. This strongly supports our main conclusion that these divergent responders are not advancing as rapidly as the spring-only responders because their vernalization sensitivities are largely compensating for the later seasonal warming that would be expected to advance FFD.

In our main analysis, we used a relatively relaxed $P \leq 0.10$ significance threshold for inclusion of climate predictors in our regression models. Our model-building procedure involved fitting a regression model with up to two β-parameters, and we had concerns regarding statistical power because of (*i*) the relatively small sample size of our time series (median $n = 37$ for Chinnor and median $n = 32$ for Washington, DC) and (*ii*) the high variance and noise in the phenological observations. Regarding point *ii*, we expected additional variance not accounted for in our models because factors other than temperature sums may also affect flowering times in some of our sampled species, a point we mention in our discussion. These other factors may include, for example, other climatic influences (e.g., moisture, radiation) or issues related to the observations themselves (e.g., observer biases, recording of first rather than mean flowering date). Indeed, the interquartile range in R^2 values for the significant climate responders at Chinnor is 0.22–0.45, indicating that although our temperature sum models can explain a significant portion of the underlying variance in FFD, there is still a substantial fraction of the variance in flowering dates attributable to factors other than temperatures sums.

To show that our results are relatively robust to the significance level chosen, we have repeated our main analysis using a $P \leq 0.05$ significance threshold for inclusion of the climate sensitivities (spring warming and vernalization). This criterion selects 37 divergent responders (Fig. S1) from the available species pool and also increases the number of species that qualify as nonresponders. Repeating our trend analysis using only these 37 divergent responders yields identical results to our previous analyses. Modeled trends (Fig. S2) in the divergent responders using the spring warming predictor only are significantly different from the observed trends (Student's t test and WR test, $P \leq 0.05$), whereas trends from models using both the spring warming and vernalization predictors are not significantly different from the observations. Results and conclusions from our main analysis are therefore generally robust to our choice of significance threshold.

Our analyses are also robust to the choice of baseline for the growing degree day calculations. Using a baseline of 5 °C (compared with 0 °C in the original analysis), we identify slightly fewer divergent responders ($n = 67$) (Fig. S3). Repeating the trends analysis (comparing models with spring warming only vs.

models with spring warming and vernalization predictors) also gives nearly identical results (Fig. S4): Trend distributions were significantly different ($P \leq 0.05$) from the observations when using a spring warming model only (Student's t test and WR test) but not when using a model that incorporated both spring warming and the vernalization predictor.

Analysis: Washington, DC. For Washington, DC, we identified 77 (72.6%) spring-only responders and 11 (10.4%) divergent responders (P value method). Our alternative AIC selection criteria identified only slightly fewer divergent responders [$n = 10$ (9.4%)]. The small absolute number of divergent responders in the Washington, DC dataset makes it difficult to perform the same statistical analyses we conducted at Chinnor. Still, it is worth noting that mean spring warming sensitivities are -5.0 d·Z^{-1} and -8.4 d·Z^{-1} for the spring-only and divergent responders, respectively, indicating that the divergent responders are likely to be as or more sensitive to spring warming than the spring-only species. Differences between Chinnor and Washington, DC could arise from a variety of issues that are difficult to tease apart. Differences in divergent responders between the two sites may reflect the differential climate sensitivities across habitats and communities. Washington, DC is a warmer site, on average (mean annual temperature = 14.2 °C), than Chinnor (mean annual temperature = 10.0 °C), and species at Washington, DC may be less sensitive to

warming and chilling or may be responding to other environmental cues. Alternatively, differences in the construction of the datasets may affect our results. For example, observations from Washington, DC come from multiple observers and are drawn from a large geographic area. In contrast, the Chinnor observations come primarily from one observer who recorded observations from a more localized area and habitat (2). The Chinnor data are then likely to be more homogeneous and community-focused, and are more likely to capture the pervasiveness of vernalization at the plant community level accurately. Another possibility is power issues attributable to the fewer number of species and fewer total observations at Washington, DC. As an example, if we repeat the analysis on Chinnor but eliminate all observations before 1960, our P value selection criteria only identifies 59 divergent responders, down from 70 in the original analysis.

Species' Climate Sensitivities. As separate supporting information, we have included two tables summarizing the climate sensitivity information for Chinnor (Dataset S1) and Washington, DC (Dataset S2). Columns are genus, species, estimated vernalization sensitivity (β-coefficient), SE in the vernalization sensitivity estimate, estimated spring warming sensitivity (β-coefficient), SE in the spring warming sensitivity estimate, and the model chosen based on the $P \leq 0.10$ model selection criteria.

1. Gotelli NJ, Ellison AM (2004) *A Primer of Ecological Statistics* (Sinauer, Sunderland, MA).

2. Fitter A, Fitter R, Harris I, Williamson M (1995) Relationships between first flowering date and temperature in the flora of a locality in central England. *Funct Ecol* 9:55–60.
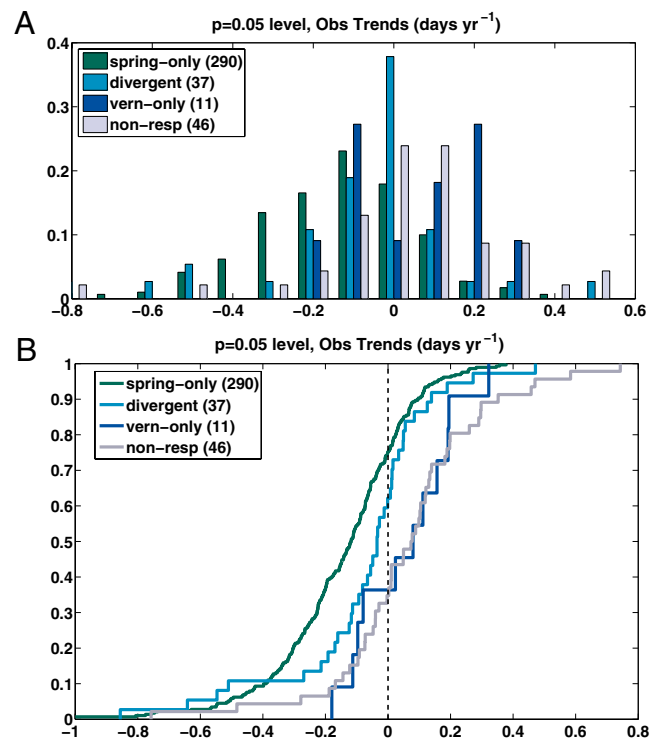
Fig. S1. Change (temporal trend) in FFD within four response categories at Chinnor, based on parameter selection using a $P \leq 0.05$ significance threshold. Plant species at Chinnor fall into four categories of observed seasonal temperature responses that covary with the species' long-term trends in FFD phenology over time: species that respond to spring warming only (spring-only, $n = 290$ species), species with both significant spring warming and fall/winter vernalization sensitivity (divergent, $n = 37$ species), species with fall/winter vernalization sensitivity only (vern-only, $n = 11$ species), and species with no significant climate sensitivity (non-resp, $n = 46$ species). Observed changes in FFD over time (x axis, trend in d·y^{-1}) are negative for species that have advanced their spring FFD and positive for species that have delayed their spring FFD. (*A*) Normalized histogram for each response category. (*B*) Empirical cumulative distribution function for each response category.

Fig. S2. Observed and modeled FFD trends (d·y^{-1}) for the 37 divergent responders at Chinnor, based on parameter selection using a $P \leq 0.05$ significance threshold. Observed trends in FFD are colored gray, modeled FFD trends using only spring warming sensitivities are colored green, and modeled FFD trends using spring warming and vernalization sensitivities are colored blue ($n = 37$ species). The observed trends are centered on zero, and trends for the models incorporating vernalization sensitivity are similarly centered. If only the spring warming predictors are considered for these species, modeled trends are biased negative, predicting a general advance in the timing of FFD that is at odds with the observations. Including vernalization sensitivity as well as spring warming sensitivity improves the ability of the model to hindcast, and match, the observed temporal FFD trends.
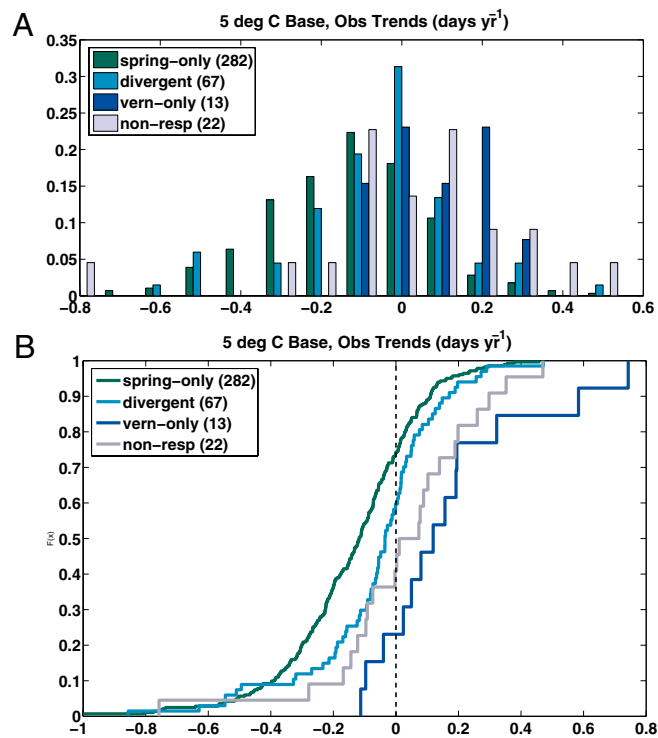
Fig. S3. Change (temporal trend) in FFD within four response categories at Chinnor, using a 5 °C baseline for the growing degree day calculations. Plant species at Chinnor fall into four categories of observed seasonal temperature responses that covary with the species' long-term trends in FFD phenology over time: species that respond to spring warming only (spring-only, $n = 282$ species), species with both significant spring warming and fall/winter vernalization sensitivity (divergent, $n = 67$ species), species with fall/winter vernalization sensitivity only (vern-only, $n = 13$ species), and species with no significant climate sensitivity (non-resp, $n = 22$ species). Observed changes in FFD over time (x axis, trend in d·y^{-1}) are negative for species that have advanced their spring FFD and positive for species that have delayed their spring FFD. (A) Normalized histogram for each response category. (B) Empirical cumulative distribution function for each response category.
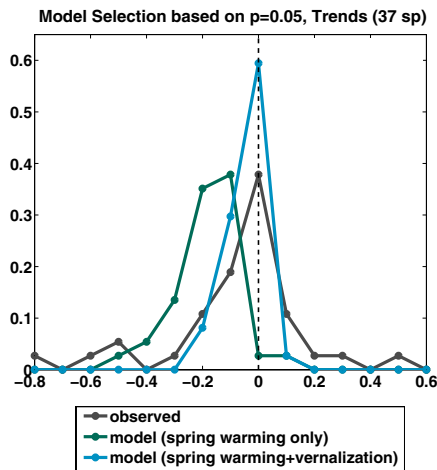
Fig. S4. Observed and modeled FFD trends (d·y⁻¹) for the 67 divergent responders at Chinnor, using a 5 °C baseline for the growing degree day calculations. Observed trends in FFD are colored gray, modeled FFD trends using only spring warming sensitivities are colored green, and modeled FFD trends using spring warming and vernalization sensitivities are colored blue (*n* = 67 species). The observed trends are centered on zero, and trends for the models incorporating vernalization sensitivity are similarly centered. If only the spring warming predictors are considered for these species, modeled trends are biased negative, predicting a general advance in the timing of FFD that is at odds with the observations. Including vernalization sensitivity as well as spring warming sensitivity improves the ability of the model to hindcast, and match, the observed temporal FFD trends.


Fig. S5. Observed and modeled FFD trends (d·y⁻¹) for the spring-only responders at Chinnor. Observed FFD trends are colored gray, and modeled FFD trends using estimated spring warming sensitivities are colored green. Negative values indicate advancement of spring FFD, and positive values indicate delay of spring FFD. (*A*) *P* value-selected species (*n* = 275 species). (*B*) AIC-selected species (*n* = 298 species). Trends in *A* and *B* are not significantly different by the Student's *t* test but are significantly different by the WR test. The WR test finds a significant difference in the right tails of the distributions, where the model fails to predict the minority of species (about 15%) that exhibited delayed responses.

Table S1. Observed FFD trend comparison: Spring only vs. divergent responders at Chinnor

Selection criteria	Tests for differences in observed FFD trends between spring-only vs. divergent responders by:		
	t test	KS test	WR test
P value (70 sp)	0.020	0.002	0.002
AICc (47 sp)	0.150	0.049	0.035

Significance (*P* value) of various statistical tests comparing observed FFD trends between the spring-only and divergent responders: two-sample Student's *t* test (*t* test), two-sample KS test, and WR test. Results for both divergent model selection criteria (*P* value method and AICc method) are shown. sp, species.

Table S2. Spring warming sensitivity comparison: Spring only vs. divergent responders at Chinnor

Selection criteria	Tests for differences in strength of spring warming sensitivities between spring-only vs. divergent responders by:		
	t test	KS test	WR test
P value (70 sp)	0.593	0.816	0.935
AICc (47 sp)	0.134	0.193	0.114

Same as Table S1 but comparing spring warming sensitivities. Spring-only species have significant spring warming sensitivities only. Divergent species have both spring warming and fall/winter vernalization sensitivities. sp, species.

Table S3. Observed vs. model trend comparison: Spring-only responders at Chinnor

Selection criteria	Tests for differences in observed vs. modeled FFD trends by:	
	t test	WR test
P value (275 sp)	0.303	0.026
AICc (298 sp)	0.158	0.008

Same as Table S1 but comparing modeled vs. observed trends for the spring-only responders. Individual species models are based on the singular warming sensitivities derived via our model selection procedure. sp, species.

Table S4. Observed vs. model trend comparison: Divergent responders at Chinnor

Selection criteria	t test	WR test
P value (spring warming only)	0.007	≤0.001
P value (spring warming + vernalization)	0.374	0.9154
AICc (spring warming only)	0.024	≤0.001
AICc (spring warming + vernalization)	0.315	0.892

Same as Table S3 but comparing modeled vs. observed trends for the divergent responders. Spring warming only refers to species models using only spring warming sensitivities; spring warming + vernalization refers to species models refit to incorporate both vernalization and spring warming sensitivities.

Other Supporting Information Files