Evolved phenological cueing strategies show variable responses to climate change

Running title: Evolved cue use leads to variable shifts

Keywords: phenological shifts, climate change, evolved cueing strategies, cryptic genetic variation, novel climates, cue integration

Article type: Article

Article length: 247 words in the abstract, 7716 words in the main text, nine figures, six supporting figures, supporting methods and a supporting table

Abstract

Several studies have documented a global pattern of phenological advancement across multiple taxa that is consistent with ongoing climate change. However, the magnitude of these phenological shifts is highly variable across taxa and locations. This variability of phenological responses disrupts species interactions under climate change, but has been difficult to explain mechanistically. To understand how climate change can create variable responses in different groups of organisms, we constructed a model for the evolution of phenological cueing strategies using historic climate data from 78 locations in North America and Hawaii. In this model, organisms evolve strategies that integrate multiple environmental cues to inform anticipatory phenological decisions. Across locations, organisms in our model evolved diverse strategies that were spatially autocorrelated on a continental scale and partially predicted by hypothesized explanatory factors, showing that similar strategies tend to evolve in similar climates. Within locations, a wide range of evolved strategies showed similar response phenotypes and fitness outcomes under historical conditions. However, these strategies responded differently to novel climatic conditions, with variable fitness consequences. These cryptic differences in cueing strategies evolved under historical conditions because epistasis and non-additive genotype × environment interactions among years resulted in extensive but constrained regions of trait space with similar fitness. These findings show how the evolution of phenological cueing strategies can explain observed variation in phenological shifts and unexpected responses to climate change. While phenological cueing strategies evolve in predictable ways under historical climate regimes, they still express highly variable responses to novel climate change.

Introduction

Recent years have seen increasing interest in the study of phenological shifts. While organisms around the world have generally shown a "global coherent fingerprint" of advancing phenology with climate change (Parmesan and Yohe 2003; Parmesan 2007; Thackeray et al. 2010), several studies also point to substantial unexplained variation in phenological shifts (Parmesan 2007; Thackeray et al. 2010; Pearse et al. 2017). This variation in responses to climate change is an important factor driving phenological mismatch and the disruption of species interactions (Parmesan 2006; Kharouba et al. 2018). It has become increasingly clear that understanding how organisms integrate multiple environmental cues will be necessary to explain and predict phenological shifts (Forrest and Miller-Rushing 2010; Visser et al. 2010; Pau et al. 2011; Chmura et al. 2019).

In order to examine the causes of variation in phenological shifts, we developed a general model that simulates the evolution of integrated cueing strategies, and applied it to different environmental histories. We hypothesized that organisms experiencing different environmental histories would evolve different phenological strategies, caused by consistent differences in the reliability of predictive information provided by different kinds of environmental cues (Reed et al. 2010; McNamara et al. 2011; Chevin and Lande 2015). We further hypothesized that these variable strategies create variation in phenological responses to climate change (Bonamour et al. 2019).

This model of phenological cue evolution combines three key features: 1) organisms use a weighted sum to combine inputs from multiple cues, 2) organisms make a phenological decision in response to a threshold of this weighted sum and 3) organismal sensitivity to each environmental cue is an evolved trait. These three features produce a relatively simple and interpretable model of phenological cue evolution, and have been commonly described in cue-response mechanisms across a wide range of organisms. Many organisms balance information from multiple cues to arrive at an integrated perception of their environment (Seeley 1995; Gu et al. 2008; Wilczek et al. 2010; Burghardt et al. 2014) and use

response thresholds to trigger a behavioral outcome (Beshers and Fewell 2001; Gu et al. 2008). The weighted sum structure of our model finds parallels at the behavioral and neural level in organisms (Ernst and Banks 2002; Ma et al. 2006), and these cue-response systems have been shown to evolve though changes in weighted sensitivities to cues (Seeholzer et al. 2018). More generally, our model faces the problem of separating informative signal from noise (McNicol 2005), and applies a genetic algorithm to develop a predictive multimodal cueing strategy over multiple generations; in this case, populations in our model are shaped by evolution to identify optimal phenological cueing strategies in the context of noisy and complex environments.

This model was developed to provide qualitative rather than quantitative predictions. For this reason, our aim was to develop a relatively simple and general model that would provide interpretable results. We then applied this model to real climate data from 78 locations across North America and Hawaii in order to explain observed variation in phenological responses to climate change.

Model and methods

Our model simulates the evolution of a generalized asexually-reproducing organism in a simplified environment defined by daily maximum temperature, total daily precipitation, and day of the year (hereafter, temperature, precipitation and day). These daily conditions provide cues to anticipate future environmental conditions each year, and also determine the fitness of individuals in the population (Fig. 1). We first present the decision-making model our organisms use, and then describe the fitness functions that determine the consequences of phenological decisions.

Cue integration

In our model, the environmental cues E on each day are cumulative annual daily maximum temperature (γ_{temp}), cumulative annual daily precipitation (γ_{precip}), and day of year (γ_{day}):

$$E = [\gamma_{temp}, \gamma_{precip}, \gamma_{day}] \qquad (1)$$

The use of cumulative annual temperature and precipitation is based on the assumption that organisms are aware of and can be influenced by past environmental conditions, consistent with degree-day models of development and phenology. Day of year provides a proxy for a consistent and non-climatic environmental cue akin to photoperiod, implicitly assuming an underlying physiological mechanism with which organisms are able to infer the day of the year from photoperiodic dynamics with equal accuracy across all locations. Although the amplitude of seasonal photoperiodic changes is larger at higher latitudes, this simplifying assumption is supported by studies showing that tropical species are able to detect extremely small changes in photoperiod near the equator (Hau et al. 1998; Dawson 2007). More fundamentally, this assumption also allows us to conservatively infer the relative information content of a climatically invariant cue across multiple locations, separate from the effect of increasing photoperiodic amplitude at higher latitudes. However, using cumulative photoperiod produced qualitatively similar results (e.g., Fig. S1).

Each individual in our model has a genotype (G) defined by three traits (τ), which reflect its sensitivity to the three environmental cues:

$$G = [\tau_{temp}, \tau_{precip}, \tau_{day}] \qquad (2)$$

Each day of the simulation, each individual combines its cues and genotype into a weighted sum, which represents the response sum (S):

$$S = \frac{\gamma_{temp}}{\tau_{temp}} + \frac{\gamma_{precip}}{\tau_{precip}} + \frac{\gamma_{day}}{\tau_{day}} \qquad (3)$$

When this sum exceeds the response threshold S≥1, the organism makes an irreversible phenological decision in anticipation of future fitness conditions. The genotype G then represents the inverse weights

of our weighted sum. We use $\frac{1}{\tau}$ as the weights for the response sum for interpretability and consistency; genotypic traits are represented in the same units as the cue itself, and trait values indicate the critical cue value that would trigger a phenological response in the hypothetical absence of other cues. This also means that fixed increases or decreases to trait have the same effect regardless of trait value (e.g. increasing τ_{day} by 1 means that in the absence of other cues the organism would respond one day later, regardless of whether τ_{day} was previously 3 or 300). As a consequence, large trait values correspond to low sensitivity, and low trait values to high sensitivity.

Environmental data

We used our model to simulate the evolution of phenological strategies in 78 locations across North America and Hawaii. We used historic climate records at these locations, for which we had an average of 98 years of daily temperature and precipitation data.

All available years of daily maximum temperature (degrees Celsius) and daily precipitation (mm rainfall equivalent) data were obtained from the NOAA Climate Data Online portal (Climate Data Online 2018) for 82 locations in North America and Hawaii. Years with <325 daily temperature and precipitation observations were excluded from further analysis, and four locations with <50 years of data remaining were also excluded. For each of the remaining 78 locations, the interquartile range (IQR) was calculated as the difference between first quartile (Q_1) and third quartile (Q_3) observations. Temperature observations less than (Q_1-4*IQR) and greater than (Q_3+4*IQR) were identified as extreme outliers likely resulting from measurement error, and were excluded; such outliers were a small proportion (0.0012%) of the overall dataset. Missing observations in the remaining dataset (less than 1% of observations) were imputed using an expectation-maximization with bootstrapping (EMB) algorithm. Imputed values for temperature and precipitation were bounded by the observed minimum and maximum values of each location, and informed by priors based on the means and standard deviations of each location. This

procedure imputed a complete dataset of daily maximum temperatures and daily precipitation for each location, with an average duration of 98 years (SD = 18.9 years) (see Supporting Data Table 1).

To ensure that cue values were always non-negative, the temperatures for each location were shifted so that the minimum transformed temperature for that location was zero. Environmental moisture (*m*) was calculated based on daily precipitation totals (*p*) in the dataset using a formula that includes a proportional retention constant (α, set to 0.8) to represent the partial retention of moisture over time, as well as the input of new precipitation each day (p) (equation 4). Changing the retention constant did not qualitatively change the model.

$$m_t = m_{t-1}\alpha + p_t \qquad (4)$$

Day of the year was represented as an integer value reflecting the number of days since January 1 of each year inclusive. The 366th day was truncated from leap years in the dataset.

Fitness

Individuals reproduce at a rate proportional to the sum of the daily fitness they accrue over a fixed window starting one day after exceeding their response threshold. The fitness gained on any given day is the product of two skew-normal function outputs – one based on temperature, the other on moisture. These two fitness functions are combined to yield a 2-dimensional fitness surface akin to a quantitative version of a 2-dimensional Hutchisonian niche (Hutchinson 1957). We used a skew-normal distribution because the thermal performance curves of ectotherms are generally asymmetrical, with a sharp decline above their optimal temperature and a more gradual decline below it (Huey and Stevenson 1979; Sinclair et al. 2016). For simplicity, we used the same skew normal functional form (with a skew parameter of -10) for both temperature and moisture, though this model showed qualitatively similar results with alternative fitness functions (see *Sensitivity analysis*). These temperature and moisture performance functions were parameterized separately for each location, such that the peak for each occurred at the 90th

percentile of all daily observations for a given location, and the function had a value that was 10% of the peak when the cue was at the 10th percentile of all daily observations. This parameterization assumes that both the temperature and moisture performance curves are locally adapted to the prevailing conditions at each location, with an expected 10% of temperature and moisture values exceeding the optimal values. This allowed us to interpret differential patterns of cue use in space conservatively, without assuming underlying differences in organismal performance curves across locations. To evaluate the robustness of observed results, we tested this model at a range of alternative fitness parameterizations and observed qualitatively identical results (see *Sensitivity analysis*). Because the skew normal function does not have a simple mathematical relationship between its parameters and the location of the peak, we fit the parameters of the function to our percentile parameters by minimizing the sum of squared errors. After both skew normal functions were parameterized, we calculated the daily fitness payoff for each day in each year of each climate as the product of the two skew normal function outputs.

We chose to use a fitness function that depends on the multiplicative effects of temperature and moisture for two reasons. First, organismal fitness in the real world depends on multiple environmental factors, and our cueing model already included temperature and precipitation data as widely available and qualitatively different sources of environmental information. Allowing both kinds of data to influence fitness was more realistic than having a single factor affecting fitness, without imposing an uninterpretable degree of complexity. While this model does not assume any specific biological mechanism by which these two factors influence fitness, it assumes that these two fitness factors interact multiplicatively rather than additively. Under these assumptions, favorable conditions in both dimensions are non-substitutable requirements for fitness, consistent with the Hutchisonian niche concept. Second, having two factors multiplicatively affect overall fitness prevented cue integration strategies from evolving in ways that directly mimicked the underlying fitness function to an unrealistic degree. Thus, this fitness function balanced biological realism with interpretability.

The raw fitness of each individual (W_i) was calculated as the sum of these daily fitness payoffs over a 10-day window beginning one day after the response sum exceeded the response threshold. This window represents a period when environmental conditions have a strong effect on fitness in the life history of our model organism. For example, this period could represent the entire active, non-diapause phase of an organism's life history or a period of establishment where the organism is vulnerable to environmental conditions; for simplicity, we assume that environmental conditions do not have fitness effects outside of this window. The same general patterns emerge with longer windows, but longer windows obscure the consequences of the phenological decision and reduce differences between organisms (see *Sensitivity analysis*). We set this fitness window to begin one day after the response threshold is exceeded in order to simulate a delayed developmental process that required some minimal anticipatory forecasting. Functionally, this one-day lag reduced the overall correlation between observed cues and experienced conditions. Increasing this lag makes it more difficult for strategies to anticipate future conditions, but did not qualitatively change the behavior of this model (see *Sensitivity analysis*). All organisms in this model were constrained to have one generation per year; organisms that did not respond by the end of the year received zero fitness. This constraint prevented the evolution of multiple generations per year or multi-annual life histories, allowing us to focus on the seasonal phenology of annual organisms.

Individuals reproduced asexually with mutation (see *Heritability and mutation*), with population size held constant and expected realized fitness of each individual proportional to its calculated relative fitness. Reproduction was implemented as a lottery model to allow for demographic stochasticity. For each evolved strategy in the final generation, we calculated its geometric mean raw fitness across all years of environmental conditions to represent its expected long-term relative fitness in that environment.

Heritability and mutation

Offspring were given the same genotypes as their parent, modified by mutation. We modeled mutation by adding to the parental traits small random numbers drawn from a normal distribution with mean 0 and a

small standard deviation. We set the standard deviation of mutation for each trait to be 0.5 percent of the overall cue range in order to produce mutation distributions with the same expected effect size in each location. In the case of the day cue, we used 360 as the maximum, leading to a standard deviation of 1.8 for mutation rate of the day trait in all locations. To improve computational efficiency, we assumed that each trait mutated for each individual in order to increase the adaptive potential of each simulated generation in each population.

Initialization and execution

For each location, we ran 60 simulations with the same parameters. Each simulation included 1000 years of climatic data resampled with replacement (i.e., 1000 generations), and a population of 500 individuals.

For each simulation, each individual in the initial generation was assigned uniform random trait values between 0 and 4 × (the maximum cue value in that location, or 360 in the case of the day cue). This results in an initial population of individuals with response phenotypes ranging between responding on day 1 and never responding (see Fig. 1D for an example). Each simulation run experienced 1000 years of environmental conditions drawn by year with replacement from the set of available environmental data for that location (Fig. 1 and Supporting Table 1). These draws were different for each simulation.

Assessing realized relative cue use

Trait values represent sensitivity to cues, but are location-specific; they represent hypothetical threshold cue values that would trigger a phenological response in the absence of other cues. The different ranges of cue values in different locations mean that the same trait values produce different behavior in different locations. In order to compare strategies between locations, we define the "trait effect" (T) as a metric of proportional cue use. This metric quantifies each organism's realized reliance on different cues represented by the proportion of the response sum that is contributed by each $\frac{\gamma}{\tau}$ term on the day the response threshold is exceeded. This trait effect metric is a mathematical composition (a vector that sums

to a constant – in our case, 1), which allows us to use compositional techniques and plot our data on ternary plots. For each individual in the final population at the end of each simulation, we calculated the Aitchison compositional mean (Boogaart and Tolosana-Delgado 2013) of the trait effects that would have been realized for each year of actual climate data used. This compositional mean represents the expected relative cue use of that genotype across all years of the historic environment.

Analysis of explanatory factors

We conducted analyses to examine correlations between evolved strategies and three sets of potential explanatory variables. The first set of analyses considered five location variables that provide a broad biogeographic description of each location: distance to coast, elevation, latitude, mean annual precipitation and mean annual temperature. The second set of analyses focused on six variables that quantify climatic variance and predictability: the mean annual coefficient of variation for daily maximum temperature, the mean annual coefficient of variation for daily precipitation total, the coefficient of variation for annual mean daily maximum temperatures, the coefficient of variation for annual mean daily precipitation totals, the lag=1 autocorrelation coefficient for daily maximum temperature, and the lag=1 autocorrelation coefficient for daily precipitation totals. The first two of these variables provide metrics of intra-annual climatic variation, the second two provide metrics of inter-annual climatic variation, and the last two provide metrics of short-term predictability. The third set of analyses used six published metrics of climatic predictability, variability and seasonality (Pau et al. 2011; Lisovski et al. 2017): Lisovski et al.'s predictability and seasonal amplitude (hereafter *seasonality*) metrics for temperature and precipitation, and Pau et al.'s variance metrics for temperature and precipitation. Lisovski's metric of predictability quantifies the ability of a model parameterized with a moving window of 4 years to predict climatic conditions in the following year across the time series. Pau's metric of temperature variance is the standard deviation of mean monthly temperatures, and Pau's metric of precipitation variance is the coefficient of variation for mean monthly precipitation (Pau et al. 2011). Lisovski's metric of seasonality (i.e., seasonal amplitude) was the difference between the upper and lower 2.5^{th} quantile of the annual

distribution (Lisovski et al. 2017). Because the Pau et al. and Lisovski et al. metrics of seasonality were highly correlated, we did not also include the Pau et al. metrics of seasonality.

These analyses used a dataset composed of the mean strategies that evolved in each simulation conducted in each location. For each analysis set, we used linear mixed models including all *a priori* explanatory variables as fixed factors for each trait effect dimension, with an additional random factor to allow intercepts to vary by location. For each analysis set, we ran the linear mixed model separately for each of the three cues. We find qualitatively identical results using logit-transformed trait effects, but present analyses of untransformed data here so that the effect sizes are reported in interpretable units (Fig. S3). Standardized effect sizes (B) are represented by the fixed effect slope coefficient divided by the standard deviation of each explanatory variable to allow for comparisons between explanatory factors; these effects sizes reflect the change in trait effects across one standard deviation of the explanatory variable (Bring 1994).

GxE calculations

As our model is deterministic, genotype by environment interactions explain all variation in fitness in each location that is not explained by average differences between individuals and between years. To calculate the percent of fitness variation explained by GxE effects, we calculated the fitness of the final population of each simulation when exposed to every year of climate data from that location. We then fit this data with a linear regression, with fitness as the response variable and individual and year as fixed effects. Any variation not explained by individual or year effects was generated by GxE effects, so we calculated the percent fitness variation explained by GxE effects for each simulation as $(1 - r^2) \times 100$.

Response similarity

In order to present patterns of similar response phenology (Fig. 6B), we calculated a “response similarity” metric. For each location, we calculated the fitness and response day of all genotypes in a 100x100x100

cube of trait values ranging from the 10th through 90th percentiles of observed cues in that location. We then compared the response day of each individual in each year to the response day of the overall most fit genotype (highest geometric mean fitness) in that year. Taking the absolute value of these differences and dividing by the maximum difference in that year gave the proportional response dissimilarity for that year; one minus the dissimilarity gave us the similarity. We then averaged this across all years for that location; this provides a measure of how similar each genotype's response phenology was to that of the most-fit genotype.

Climate change scenarios

We examined how the individuals from the final generation of each simulation performed in novel climate regimes using two simple climate change scenarios. In the "shift" scenario, we advanced the historic temperature and precipitation regime in each year by 5 days. In the "warming" scenario, we increased all daily temperatures by 3 degrees; the precipitation regime was unchanged. These two scenarios are not intended to represent detailed or realistic climate change scenarios; instead they reflect a simple and systematic erosion of the key correlations in historical climatic data. In the *shift* scenario, the historical correlations between climatic conditions (temperature and moisture regime) and day of year are weakened, but the key correlation between the temperature and moisture regimes is maintained. More importantly, the seasonal fitness landscape (the timeline of potential daily fitness payoffs) in any given year remains unchanged in the *shift* scenario, except that it is advanced by 5 days. In contrast, the *warming* scenario weakens historical correlations with temperature relative to the other two environmental cues, and also fundamentally changes the seasonal fitness landscape in any given year. Thus, these two novel climate scenarios both represent important departures from the historical climate regime, and also reflect qualitative differences in the severity of these departures. In both scenarios, we calculated the response day and fitness that would have been realized for each individual of our final populations in each unique year of the modified climate regime. This allowed us to assess how climate

change affected the phenotype and fitness consequences of each genotype that evolved under historical conditions.

For both climate change scenarios, we assessed correlations between each historic trait effect (T) and the change in response timing, and between each trait effect and the change in geometric mean fitness for each evolved genotype. In all cases we used linear mixed models with location as a random factor, allowing intercepts and slopes to vary. For these analyses, we report unstandardized effect sizes (β) as the slope coefficient of each fixed explanatory factor; in these analyses, the effect size presents the expected change in response day or geometric mean fitness with a one unit change in the trait effect. We use the unstandardized effect size in these analyses because the explanatory factors are all trait effects, and so are all on the same scale without standardization.

Sensitivity analyses

To ensure that the qualitative findings of our model were robust, we tested different model structure, different cues, and different parameter values. To test if our results were sensitive to model structure, we ran the simulation with both linear and quadratic terms for each of our cues. To test if our results were sensitive to specific cues used, we ran the simulation using (a) daily precipitation, daily temperature, and day of year, and (b) cumulative temperature, cumulative precipitation, and cumulative photoperiod. To test if our simulation was sensitive to the functional form of the fitness surface, we ran the simulation on a subset of locations using a product of symmetrical Gaussian distributions, and a product of skew Gaussian distributions with other skew parameter values. To test if our simulation was sensitive to parameter values, we ran the simulation on a subset of locations using larger and smaller population sizes, longer and shorter simulation durations, larger and smaller mutation distances, reduced mutation rates, fitness windows of 5, 10, and 20 days, and higher and lower moisture retention coefficients. We also tested the robustness of a key finding of our model (a broad range of genotypes leading to similar fitnesses) by calculating geometric mean fitness across a broad range of trait values for each location for

(a) base parameter values, (b) two alternate parameterizations for each location, each having randomly assigned fitness windows (5, 10, or 20), and parameters for the fitness function.

Results

Selection favored combinations of traits (i.e., genotypes) that yielded the highest geometric mean fitness. In many simulations, a region of successful trait combinations was identified relatively quickly, with selection, mutation and drift leading to gradual shifts in the average population genotype, as well as the branching and pruning of lineages over time (e.g., Fig. 2). Some simulations experienced large shifts in trait use over time, often with concurrent changes ramifying across multiple traits. Most of the individuals in the final generation emerged from the dominant evolved lineage, and showed similar combinations of traits. Thus, individual variation within each simulation was well-represented by the compositional mean strategy for that simulation, although these mean strategies showed considerable variation between simulations (Fig. S4).

We found considerable spatial variation in the average (compositional mean across simulations) evolved strategy for each location (Fig. 3). Geographic patterns in cue use differences between locations are visible to the eye when mapped (Fig. 3a), and showed strong positive spatial autocorrelation on a continental scale (Fig. 3b-d). This result indicates that similar mean strategies evolved under similar climates, suggesting some underlying predictability in the evolution of phenological cueing strategies.

Several location-based explanatory factors were associated with variation in phenological cueing strategies (Fig. 4). For example, the evolved use of day cues (measured as the mean trait effects of the final generation) declined with distance from the coast (B=-0.112), and increased with elevation (B=0.103). The use of day cues may have been favored in more moderated coastal climates if these climates tend to show more predictable seasonal variation where climatically plastic cues are less relevant for identifying successful windows of opportunity. Conversely, the greater use of day cues may have

evolved at high elevation locations if more variable day-to-day climatic conditions meant that climatic cues provided less predictive information than a climatically invariant day cue. In comparison, these two explanatory factors suggest that the evolution of phenological cues may reflect a key distinction between informative climatic variation and noise. In addition, locations with higher mean annual precipitation tended to rely on day cues less (B=-0.120), and generally showed a greater emphasis on precipitation cues (B=0.167). This pattern could result from the absence of informative cueing information provided by precipitation cues in relatively dry locations.

Some metrics of climatic variation also explained the evolved use of phenological cues. Locations with high mean annual CV_{precip} (i.e., high within-year precipitation variability) tended to show a reduced use of day cues (B=-0.166), and an increased use of precipitation cues (B=0.307). In contrast, locations with high CV of mean annual precipitation (i.e., high between-year precipitation variability) tended to show a reduced use of precipitation cues (B=-0.167). These patterns suggest that the use of precipitation cues allows for adaptive plasticity in locations with variable within-year precipitation regimes, but a reliance on precipitation cues is not favored in locations with high between-year variation in precipitation, where precipitation cues may be lacking or uninformative in some years. Unexpectedly, locations with high lag 1 autocorrelation in temperature (ACF_{temp}) actually tended to use temperature cues less (B=-0.164), while locations with high ACF_{precip} tended to use precipitation cues less (B=-0.130), and day cues more (B=-0.123). This result is counterintuitive, since it shows a negative correlation between a direct measure of (daily-scale) climatic predictability and evolved cue use. However, a similar pattern emerges from an independent published metric of predictability on a seasonal scale: locations with high Lisovski temperature predictability showed a reduced use of temperature cues (B=-0.173), and locations with high Lisovski precipitation predictability showed a reduced use of precipitation cues (B=-0.141). We speculate that these patterns occur because the use of day cues can compensate to some degree for the use of climatic cues in locations with highly predictable climatic conditions, and because maximizing the fitness outcomes of strategies in locations with one highly predictable climatic condition may depend more on

identifying windows of opportunity constrained by the other climatic factor. Pau's metric of precipitation variance was associated with an increased use of day (B=0.137) and temperature (B=0.143) cues, and a reduced use of precipitation cues (B=-0.228), while Lisovski's metric of precipitation seasonality was associated with a reduced use of day (B=-0.176) and temperature (B=-0.107) cues, and an increased use of precipitation cues (B=0.280). These patterns are in the opposite direction because the Pau measure of precipitation variance uses the coefficient of variation, and so is negatively correlated with mean annual precipitation (ρ=-0.54, p<0.0001)), while Lisovski's metric of precipitation seasonality was highly positively correlated with mean annual precipitation (ρ=0.94, p<0.0001). The pattern with Lisovski's measure of precipitation seasonality supports the previous observation that the use of precipitation cues is positively correlated with intra-annual precipitation variation. Overall, while several factors emerged as meaningful predictors of phenological cue use in these analyses, most of the variation in cue use was unexplained even in models that combined all 17 factors for each trait effect (T_{day}: marginal R^2=0.102; T_{temp}: marginal R^2=0.110; T_{precip}: marginal R^2=0.308).

Comparing across simulations, the mean genotypes that evolved from repeated runs in the same location often showed highly variable phenotypic cueing strategies. This variation in strategies can be observed in the ternary plots from each location (Fig. 5, Fig. S2), which show the mean strategies (represented as mean realized trait effect combinations) that evolved across 60 simulations in each location. While some locations show highly clustered strategies, most locations show strategies arrayed along a single dimension, or across a broad range of evolved strategies. Notably, despite this variation in phenological cueing strategy, many of these strategies showed geometric mean fitnesses that were similar to the most fit mean genotype across all simulations (Fig. 5, Fig. S2, Fig. S4). Most locations showed a substantial diversity of successful strategies; in the ternary plots for each location (Fig. 5, Fig. S2), gray convex hulls include all simulations where the mean evolved strategy was within 10% of the most fit mean genotype across all simulations. Viewed another way, the range of observed variation in traits for each location was much larger than the range of observed variation in geometric mean fitness outcomes across all

simulations in each location (Fig. 6). Figure 6 shows that diverse strategies across a wide range of trait effects yield small variation in fitness outcomes. These results indicate that simulated evolution in our model found a wide range of alternate strategies that achieved similar fitness outcomes.

An examination of the geometric mean fitness for many combinations of trait values shows a clear high-performance fitness volume that spans a wide range of trait values (Fig. 7A). Fitness within that volume is nearly constant, while strategies outside of that region have dramatically reduced fitness. The similarity of fitness outcomes within that region can be explained by examining the average similarity of the phenological response phenotypes (i.e., day of response) (Fig. 7B): genotypes along that high-fitness surface show similar response phenotypes under the historic climate, demonstrating that this volume represents a set of strategies within which a wide range of trait combinations can be integrated to yield similar phenological behavior and fitness outcomes under the historical climate regime.

The effect of genotype was inconsistent across years when evaluating the final set of genotypes that evolved from each simulation (Fig. 8). This genotype x environment interaction accounted for 29% of fitness variation across all locations, and reduced the consistency of fitness advantages between genotypes.

In the *shift* climate change scenario, populations advanced their phenology, but showed highly variable changes in their realized fitness (Fig. 9a). These effects were non-random; as expected, organisms that relied more on day cues were less likely to advance their phenology on pace with the changed climate (β=4.6, $p<0.0001$), while organisms that relied more on temperature or precipitation cues were more likely to advance their phenology (temperature, β=-2.5, $p<0.0001$; precipitation, β=-2.2, $p<0.0001$). Because the seasonal fitness landscape retained the same shape but advanced by 5 days in this scenario, organisms that relied more on day cues generally showed a weak pattern of more negative fitness consequences (β=-.00023, p=0.0009, Fig. 9a), while those that relied more heavily on temperature or precipitation cues showed weakly positive fitness consequences (temperature, β=0.00017, p=0.002;

precipitation, β=0.00009, p=0.04). Here it is worth noting that the coefficients are showing units of change in raw fitness (which due to the nature of our fitness functions are necessarily much smaller than 1) rather than changes average realized offspring; for comparison, the geometric mean raw fitness across all final individuals in the historic climates was 0.0098. There results were broadly consistent with our *a priori* expectations based on the coordinated shift of both temperature and precipitation regimes in this scenario; in this scenario, the seasonal fitness landscape is unchanged (albeit advanced) because both of the climatic factors that contribute to it are shifted with perfect coordination. However, even under the relatively simple shift scenario, the changes in the mean fitness outcomes for each location were unexpectedly complex, and variable. While most locations experienced a reduced mean fitness under the changed climate, some locations showed higher overall fitness (Fig. 9). The behavior of individual mean genotypes within each location shows even greater variation in response to climate change, with small changes in phenological response phenotype driving unexpectedly large changes in mean fitness outcomes under the shift scenario (Fig. S6).

Under the *warming* scenario, mean strategies with greater reliance on day or precipitation cues also showed reduced phenological advancement (day, β=8.1, p<0.0001; precipitation, β=7.7, p<0.00001), while those that relied more on temperature showed greater phenological advancement (temperature, β=-15.4, p<0.00001). However, in the warming scenario, organisms with greater reliance on day and precipitation showed a weak and unexpected pattern of higher fitness (day, β=0.0006, p=0.0026; precipitation, β=0.0007, p=0.01), while those that relied more on temperature showed a weak pattern of reduced fitness (β=-0.001, p<0.00001). In this scenario, the fitness consequences of phenological advancement were not consistently positive. Whereas the shift scenario maintained historic correlations between temperature and precipitation regimes, the warming scenario increased temperatures independently of precipitation; because fitness was a function of temperature and precipitation, this meant that the warming scenario produced novel seasonal fitness surfaces. As a result, many locations showed much larger and more severe changes in mean fitness outcomes under the warming scenario (Fig. 9b). In

some locations, realized mean fitness dropped precipitously, and these fitness effects were less predictable based on mean strategy differences across locations than in the shift scenario (Fig. 9).

Our findings were qualitatively robust across model variants that used different cues (daily precipitation, daily temperature, photoperiod, quadratic measures of cues), fitness functions, and fitness windows. Using different cues led to qualitative similar spatial patterns (e.g, Fig. S1). Different fitness window lengths (obtaining fitness for 5, 10, or 20 days after the response threshold is exceeded) produced qualitatively similar patterns, although longer windows had more consistent patterns between years, but generally reduced the fitness differences resulting from different phenological decisions (e.g., responding a day earlier or later had a smaller proportional impact on the sum of daily fitnesses). Changing mutation rates, distances, simulation duration, and population size influenced the total amount of trait change that occurred within each simulation (and in the case of duration and size, dramatically changed computational runtime), but produced qualitatively similarly patterns. Changes to fitness parameters did not change the shape of the high-fitness regions in traitspace shown in Fig. 7.

Discussion

This model suggests two key findings. First, we see that the mean evolved strategies in different locations are spatially autocorrelated (Fig. 3) and partially predictable based on location and climatic variables (Fig. 4). This shows that similar strategies evolved in locations with similar climates, and suggests that the evolution of phenological cueing strategies in our model was shaped by their environmental history in fundamentally predictable ways. We interpret these continental scale patterns in cue use as a reflection of underlying geographic patterns in the relative reliability of temperature, precipitation and day cues for predicting future fitness outcomes; in a sense, our model provides a lens with which to view patterns in real-world climatic data. In our model, selection favored cues based on their ability to predict future environmental conditions – both the ability to consistently trigger a phenological response ahead of favorable conditions, and the ability to avoid triggering a phenological response ahead of unfavorable

conditions. Even when implemented in a simple model, this selection criterion drove the evolution of phenological cueing strategies in unexpectedly complex and occasionally counterintuitive ways. Despite this, several hypothesized explanatory factors were associated with variation in cue use, including both simple location variables (including distance from the coast, elevation and mean annual precipitation) and more specific climatic variability and predictability metrics (including mean annual CV of precipitation, the autocorrelation coefficients for temperature and precipitation, and several previously published metrics of climatic variance, seasonality and predictability). While we did find factors that were significant predictors of cue use, and many of these patterns could be reasonably explained *a posteriori*, many other factors showed relatively small, inconsistent or counterintuitive effects, and we did not find unequivocal support for a consistent or intuitive relationship between the use of climatically invariant cues and simple measurements of within- or between-year climatic variation.

We suggest that the limited success of these explanatory factors for predicting cue use occurs because they are too far removed from the specific selection criterion implicit in the model. For example, metrics that attempt to summarize inter- or intra-annual climatic variation are only distantly correlated with the predictability of cues in the most relevant period of the year when fitness is highest. Likewise, even metrics that specifically target climatic predictability (as opposed to variation) were generally unable to provide a strong explanation for evolved cues because they quantify mean predictability across the year, while selection favors strategies that are able to identify potential windows of opportunity in specific parts of each year. In general, the evolved strategies in our model offers only partial support for conventional expectations based on location characteristics or standard metrics of climatic variability (Sunday et al. 2011; Molina-Montenegro and Naya 2012); instead, they illustrated the complexity of the climatic regimes and seasonal fitness landscapes that affect organisms in this model. Further, the considerable variation in evolved strategies within many locations demonstrates that climatic and environmental factors alone cannot fully predict the evolution of phenological cueing strategies.

However, the results of this analysis do suggest several potentially general patterns. Even when the specific relationships are complex or counterintuitive, the observation that some location and climatic variables are predictably associated with reliance on specific phenological cues suggests that these evolved phenological cueing strategies were a predictable outcome of their environmental histories. Moreover, evolution appeared to favor cues that were reliably present from year to year, but showed an intermediate level of variation within year – neither so predictable that they could be replaced by climatically invariant day of year cues, nor so variable that they diluted informative signal with uninformative noise. Thus, the spatial autocorrelation of evolved strategies suggests that past environmental regimes shaped the evolution of phenological cueing strategies in fundamentally predictable ways, and these analyses suggest some of the specific environmental factors that favor the different cueing strategies.

The second key finding of this model is that repeated simulations from the same location produced diverse strategies that expressed similar phenotypes under historical conditions (Figs. 6, 7), but showed strong phenotypic and fitness differences under simulated climate change (Figs. 9, S6). This mirrors the findings of empirical studies, which have shown that these differences in cue use exist between interacting species, even among those that interact strongly and often show phenological synchrony (e.g., Iler et al. 2013). The diversity of strategies that evolved in this model emerges as a result of two key mechanisms: interactions between traits in cue integration, and non-additive genotype × environment (G×E) interactions. At a fundamental level, cue integration means that multiple cues are combined to yield a single response. As a result, multiple combinations of traits can yield the same expressed phenotype. This occurs because changes in sensitivity to one cue can often be compensated for though changes in the sensitivity to another, especially in the context of climatic cues. Mathematically, the cue integration function used in our model is “non-injective”, meaning that it does not preserve a one-to-one mapping between genotypes (i.e., combinations of sensitivity traits) and expressed phenology in each year; instead, there are often multiple genotypes that can yield very similar or identical phenological

behaviors in a given year. Biologically, this situation is akin to epistasis, where interactions between traits modify their effects on the resulting response phenotype (Phillips 2008). As an extreme example, a trait that confers a very high degree of sensitivity to one cue effectively nullifies the effects of the other traits, because even a small cue value will cause the organism to exceed its response threshold. More generally, a range of genotypes can yield the same phenological behavior in any given year by differentially weighting multiple cues; this allows a wide range of strategies to show similar response dates and fitness outcomes under the same environmental conditions (Fig. 5). This can be seen when looking across three-dimensional trait space, where a large volume of genotypes across wide range of trait values show similarly high levels of expected geometric mean fitness outcomes (Fig. 7A), driven by generally similar phenological behavior under the historical climate regime (Fig 7B). In addition to epistatic interactions between traits, non-additive genotype × environment (G×E) interactions provide an additional mechanism that could contribute to the maintenance of strategy variation. The genotypes in our model show different reaction norms across years, driven by between-year differences in the historical climatic data (Fig. 8a). These G×E interactions reduce the likelihood and magnitude of consistent fitness differences between genotypes, which allows for the increased coexistence of genotype diversity. In combination, these two processes (epistatic interactions among traits and G×E interactions among years) ensure that a wide range of strategies produce similar fitness outcomes.

When exposed to climate change, the strategies that evolved in each location showed unexpectedly variable phenotypic and fitness consequences. In the *shift* scenario, organisms that were more sensitive to climatic cues showed greater phenological advancement and more positive fitness consequences, while those that relied more heavily on climatically invariant day-of-year cues showed reduced advancement and more negative fitness consequences. This result is consistent with expectations about the costs of relying on invariant day-of-year cues under a climate change scenario (Coppack and Pulido 2004; Way and Montgomery 2015), and the potential for adaptive plasticity in response to changing climates. However, in the *warming* scenario, greater reliance on the invariant day-of-year cue was generally

favorable, while organisms that relied more on temperature actually showed reduced fitness. This unexpected result occurs because the warming scenario broke the historic correlation structure between temperature- and precipitation-based factors. The two scenarios that we evaluated in this analysis differed in their effect on the seasonal fitness landscape in a fundamental way, and this difference drove the increased unpredictability of responses we observed under the *warming* scenario. Even a relatively small decoupling of the historical temperature and precipitation regimes qualitatively changed the seasonal fitness landscapes in this scenario, instead of simply shifting them earlier in time. This caused many strategies to demonstrate maladaptive plasticity in a novel climate, caused by changes in the complex structure of correlations between environmental cues and conditions. Maladaptive plasticity has been anticipated as a consequence of novel climates (Reed et al. 2010; McNamara et al. 2011; Chevin and Lande 2015; Duputié et al. 2015), and these results demonstrate the potential maladaptive responses to climate change that can occur when organisms face environments that differ from those they experienced in their evolutionary history. Thus, while intuition suggests that a greater reliance on climatic cues (as opposed to climatically invariant cues) would allow for more adaptively plastic responses to a changing climate, our findings demonstrate that this is not necessarily the case.

Our model suggests that cryptic genetic variation may result from the evolution of phenological cueing strategies. Cryptic genetic variation is genetic variation that is not normally expressed, but which can yield phenotypic variation under changed conditions (Gibson and Dworkin 2004; Gibson and Reed 2008; McGuigan and Sgrò 2009; Paaby and Rockman 2014). In the context of our model, several mechanisms fundamental to the evolution of phenological cueing strategies promote the maintenance of genotypic variation in cueing strategies which have relatively small phenotypic and fitness consequences under historic climate conditions. However, these genotypic differences in cue sensitivity can have much stronger effects on organismal responses to novel environmental regimes under climate change. Thus, the cryptic genetic variation which results from the evolution of phenological cueing strategies could contribute to observed variation in phenological responses to climate change more generally.

Our model makes several simplifying assumptions, and omits features found in many real organisms (Chuine and Régnière 2017). In particular, it would be interesting to consider a variant of this model that included sexual recombination of traits, which would likely increase standing genetic variation, and to consider more complex cue integration mechanisms, some of which could provide greater resilience under climate change. However, the qualitative conclusions of our analysis are unlikely to be altered by these omissions. The two key findings we report here have been robust across a range of model variants, and seem to emerge as a fundamental consequence of cue integration. We suggest that similar processes are likely to occur whenever multiple cues are combined to inform phenological decisions, and that this condition is likely to apply to many organisms in a wide range of contexts (Beshers and Fewell 2001; Wilczek et al. 2010; Seeholzer et al. 2018; Chmura et al. 2019).

The two key findings that emerge from this model reflect two conclusions that initially seem to be at odds. First, the finding that similar phenological cueing strategies tend to evolve in similar climates and in response to specific environmental factors suggests that environmental history shapes the evolution of phenological cueing strategies in fundamentally predictable ways. However, the second finding that evolved phenological cueing strategies show variable - and often seemingly unexpected – responses to climate change suggests that our ability to predict the plastic responses and fitness consequences of organisms living under novel climatic regimes is profoundly limited. Previous studies have anticipated aspects of these findings (e.g., Ghalambor et al. 2007; Chevin et al. 2010; McNamara et al. 2011; Michel et al. 2014; Duputié et al. 2015; Chuine and Régnière 2017), though typically without the explicit consideration of multiple cue integration. We find that while environmental factors constrain possible phenological cuing strategies, epistatic and GxE interactions often provide a wide set of successful strategies within those constraints, whose diversity is only apparent under novel climate conditions. Thus, these two conclusions both emerge from common and fundamental mechanisms that are likely to be broadly shared among diverse models of cue integration. Thus, while an examination of past environmental history could improve our understanding of phenological cueing strategies and provide a

mechanistic way to predict future phenological shifts, the nature of cue integration may put fundamental limits on our ability to predict the responses and fitness outcomes of organisms living under novel climatic regimes.

References

Beshers, S. N., and J. H. Fewell. 2001. Models of Division of Labor in Social Insects. Annual Review of Entomology 46:413–440.

Bonamour, S., L.-M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in response to climate change: the importance of cue variation.

Boogaart, K. G. van den, and R. Tolosana-Delgado. 2013. Analyzing Compositional Data with R. Springer Science & Business Media.

Bring, J. 1994. How to Standardize Regression Coefficients. The American Statistician 48:209–213.

Burghardt, L. T., C. J. E. Metcalf, A. M. Wilczek, J. Schmitt, and K. Donohue. 2014. Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. The American Naturalist 185:212–227.

Chevin, L.-M., and R. Lande. 2015. Evolution of environmental cues for phenotypic plasticity. Evolution 69:2767–2775.

Chevin, L.-M., R. Lande, and G. M. Mace. 2010. Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. PLOS Biology 8:e1000357.

Chmura, H. E., H. M. Kharouba, J. Ashander, S. M. Ehlman, E. B. Rivest, and L. H. Yang. 2019. The mechanisms of phenology: the patterns and processes of phenological shifts. Ecological Monographs 89:Early View.

Chuine, I., and J. Régnière. 2017. Process-based models of phenology for plants and animals. Annual Review of Ecology, Evolution, and Systematics 48:159–182.

Climate Data Online. 2018. https://www.ncdc.noaa.gov/cdo-web/

Coppack, T., and F. Pulido. 2004. Photoperiodic Response and the Adaptability of Avian Life Cycles to Environmental Change. Pages 131–150 *in*Advances in Ecological Research, Birds and Climate Change (Vol. 35). Academic Press.

Dawson, A. 2007. Seasonality in a temperate zone bird can be entrained by near equatorial photoperiods. Proceedings of the Royal Society of London B: Biological Sciences 274:721–725.

Duputié, A., A. Rutschmann, O. Ronce, and I. Chuine. 2015. Phenological plasticity will not help all species adapt to climate change. Global change biology 21:3062–3073.

Ernst, M. O., and M. S. Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Forrest, J., and A. J. Miller-Rushing. 2010. Toward a synthetic understanding of the role of phenology in ecology and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 365:3101–3112.

Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394–407.

Gibson, G., and I. Dworkin. 2004. Uncovering cryptic genetic variation. Nature Reviews Genetics 5:681–690.

Gibson, G., and L. K. Reed. 2008. Cryptic genetic variation. Current Biology 18:R989–R990.

Gu, Y., D. E. Angelaki, and G. C. DeAngelis. 2008. Neural correlates of multisensory cue integration in macaque MSTd. Nature Neuroscience 11:1201–1210.

Hau, M., M. Wikelski, and J. C. Wingfield. 1998. A neotropical forest bird can measure the slight changes in tropical photoperiod. Proceedings of the Royal Society B: Biological Sciences 265:89–95.

Huey, R. B., and R. D. Stevenson. 1979. Integrating Thermal Physiology and Ecology of Ectotherms: A Discussion of Approaches. American Zoologist 19:357–366.

Hutchinson, G. E. 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

Iler, A. M., D. W. Inouye, T. T. Høye, A. J. Miller-Rushing, L. A. Burkle, and E. B. Johnston. 2013. Maintenance of temporal synchrony between syrphid flies and floral resources despite differential phenological responses to climate. Global Change Biology 19:2348–2359.

Kharouba, H. M., J. Ehrlén, A. Gelman, K. Bolmgren, J. M. Allen, S. E. Travers, and E. M. Wolkovich. 2018. Global shifts in the phenological synchrony of species interactions over recent decades. Proceedings of the National Academy of Sciences 201714511.

Lisovski, S., M. Ramenofsky, and J. C. Wingfield. 2017. Defining the degree of seasonality and its significance for future research. Integrative and Comparative Biology icx040–icx040.

Ma, W. J., J. M. Beck, P. E. Latham, and A. Pouget. 2006. Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432–1438.

McGuigan, K., and C. M. Sgrò. 2009. Evolutionary consequences of cryptic genetic variation. Trends in Ecology & Evolution 24:305–311.

McNamara, J. M., Z. Barta, M. Klaassen, and S. Bauer. 2011. Cues and the optimal timing of activities under environmental changes. Ecology Letters 14:1183–1190.

McNicol, D. 2005. A primer of signal detection theory. Psychology Press.

Michel, M. J., L.-M. Chevin, and J. H. Knouft. 2014. Evolution of Phenotype–Environment Associations by Genetic Responses to Selection and Phenotypic Plasticity in a Temporally Autocorrelated Environment. Evolution 68:1374–1384.

Molina-Montenegro, M. A., and D. E. Naya. 2012. Latitudinal patterns in phenotypic plasticity and fitness-related traits: assessing the climatic variability hypothesis (CVH) with an invasive plant species. PLOS ONE 7:e47620.

Paaby, A. B., and M. V. Rockman. 2014. Cryptic genetic variation: evolution's hidden substrate. Nature Reviews Genetics 15:247–258.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology Evolution and Systematics 37:637–669.

Parmesan, C. 2007. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. Global Change Biology 13:1860–1872.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42.

Pau, S., E. M. Wolkovich, B. I. Cook, T. J. Davies, N. J. B. Kraft, K. Bolmgren, J. L. Betancourt, et al. 2011. Predicting phenology by integrating ecology, evolution and climate science. Global Change Biology 17:3633–3643.

Pearse, W. D., C. C. Davis, D. W. Inouye, R. B. Primack, and T. J. Davies. 2017. A statistical estimator for determining the limits of contemporary and historic phenology. Nature Ecology & Evolution 1:1876–1882.

Phillips, P. C. 2008. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics 9:855–867.

Reed, T. E., R. S. Waples, D. E. Schindler, J. J. Hard, and M. T. Kinnison. 2010. Phenotypic plasticity and population viability: the importance of environmental predictability. Proceedings: Biological Sciences 277:3391–3400.

Seeholzer, L. F., M. Seppo, D. L. Stern, and V. Ruta. 2018. Evolution of a central neural circuit underlies Drosophila mate preferences. Nature 559:564–569.

Seeley, T. D. 1995. The Wisdom of the Hive. Harvard University Press.

Sinclair, B. J., K. E. Marshall, M. A. Sewell, D. L. Levesque, C. S. Willett, S. Slotsbo, Y. Dong, et al. 2016. Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? Ecology Letters 19:1372–1385.

Sunday, J. M., A. E. Bates, and N. K. Dulvy. 2011. Global analysis of thermal tolerance and latitude in ectotherms. Proceedings of the Royal Society B-Biological Sciences 278:1823–1830.

Thackeray, S. J., T. H. Sparks, M. Frederiksen, S. Burthe, P. J. Bacon, J. R. Bell, M. S. Botham, et al. 2010. Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. Global Change Biology 16:3304–3313.

Visser, M. E., S. P. Caro, K. van Oers, S. V. Schaper, and B. Helm. 2010. Phenology, seasonal timing and circannual rhythms: towards a unified framework. Philosophical Transactions of the Royal Society B: Biological Sciences 365:3113–3127.

Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, performance and migration in a warming world. Plant, Cell & Environment 38:1725–1736.

Wilczek, A. M., L. T. Burghardt, A. R. Cobb, M. D. Cooper, S. M. Welch, and J. Schmitt. 2010. Genetic and physiological bases for phenological responses to current and predicted climates. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365:3129–3147.

Figure legends

Figure 1. Schematic diagram of model. a.) Genotypes combined with environmental cues (including cumulative annual daily temperature maximums, cumulative annual daily precipitation totals and day of year) result in expressed phenotypes (day of response). b.) The trait effect (T), the proportional contribution of each trait to the response decision (a representation of the interaction between genotype and environment), can be expressed as a composition and presented on a ternary plot. c.) The fitness of different phenotypes is determined by climatic (temperature and moisture) conditions during a 10-day window after the response threshold is crossed. A lottery model of reproduction determines the number of offspring produced by each individual, and mutation results in new genotypes for the next generation. d.) Selection results in evolved phenological cueing strategies that anticipate favorable conditions and avoid unfavorable conditions. In this panel, the solid blue line represents the long-term expected fitness outcome for each day under historical conditions, while the dotted black lines represent the fitness outcomes for the first and last year of the simulation for the left and right panels respectively. The black arrows at the top of each panel represent the response day of each individual of the population. Initially the timing of phenological response is spread across the year, but after 1000 generations of selection, most of the population shows similar phenological timing. This example shows the results of one simulation using climatic data from Davis, CA.

Figure 2. Temperature, precipitation and day traits evolve over 1000 generations in this representative simulation from Davis, CA. Each simulation begins with initial values for each trait drawn from broad uniform distributions; selection drives the evolution of specific strategies. The inset figure shows an expanded view of trait evolution in the final 100 generations. Each point represents the trait value of one individual; point colors show raw individual fitness proportional to the maximum value in this simulation.

Figure 3. Evolved strategies show spatial autocorrelation in relative cue use (T); similar strategies evolve in nearby locations with similar climates, and different strategies evolve in distant locations with different

climates. A) A map representing the mean evolved strategy for each of 78 locations using the compositional mean of trait effects across all years. Evolved strategies show significant positive spatial autocorrelation in reliance on B) day, C) temperature, and D) precipitation cues up to at least 1000 km. Filled circles are significantly correlated, open circles are not (Moran's I).

Figure 4. The effects of hypothesized explanatory factors on evolved cue use across all locations under historical conditions. These analyses examine A-C) simple location variables, D-F) metrics of within- and between-year climate variability, and G-I) published metrics of climatic variability, predictability and seasonality. Values indicate standardized fixed effect coefficient estimates, where the raw fixed effect coefficient is divided by the standard deviation of the predictor so that the resulting effect size represents the estimated change in trait effects with a change of one standard deviation in the predictor (see also Fig. S3).

Figure 5. Ternary plots illustrate proportional cue use at the time of response for six selected locations. Each point represents the mean strategy at the end of one simulation; each strategy is represented as a composition of the trait effects (T) in percentages, which represent relative cue use (see "assessing realized relative cue use" in Methods). Point color reflects geometric mean fitness of genotypes across all years of the climate as a percent of the maximum observed geometric mean fitness ($\bar{W}$) for each location. Simulations within 10% of the maximum observed geometric mean fitness in each location are represented as triangles and included in a gray convex hull. All other points are represented as circles. Ternary plots for all 78 locations are presented in Extended Data Fig. 1, locations are described in Extended Data Table 1.

Figure 6. Cryptic genetic variation persists under selection in part because a wide range of genotypes show similar phenotypes and fitness outcomes. In this figure, each point represents the maximum proportional range of trait effects (T) and the maximum proportional range of geometric mean fitness ($\bar{W}$) across 60 simulations for each of 78 locations. The maximum proportional range of trait effects is the

largest observed difference between the trait effects across simulations within a location, across all three trait dimensions. Likewise, the maximum proportional geometric mean fitness range is the difference between fitness of the most fit evolved strategy and the fitness of the least fit evolved strategy (i.e., equation). A conservative null expectation of a 1:1 correspondence between range in trait use and range in fitness is shown as a solid line. Strategies show a wide range of trait effects with relatively small effects on mean geometric fitness. The color of each point depicts which of the three trait effect dimensions produced the maximum range.

Figure 7. Diverse genotypes can produce similarly high fitness phenotypes. A 100x100x100 grid spanning trait values ranging from the 10th through 90th percentiles of observed cues in each location were evaluated for fitness and response day in each recorded year of climate. A.) Geometric mean fitness of each genotype is plotted in traitspace for six representative locations. Bands of high fitness span wide ranges of traits, but most combinations of traits produce low fitness. Note that (a), (c), and (d) have been rotated to place the origin in the far corner for the sake of visibility. B.) Response similarity is plotted for the same trait ranges in the same locations (see "Response similarity" in methods). Bands of high similarity span most of the traitspace, and demonstrate that many different combinations of traits can lead to similar response phenology. Comparisons with (A) show that the high fitness bands generally overlap with genotypes that respond at similar times as the highest-fitness individual, illustrating that the similarity in fitness between diverse strategies is often due to similar phenology in the historic climate regime. Because response similarity has different fitness consequences in different contexts, some genotypes with high response similarity still have low fitness (e.g. the mismatch between Barrow fitness and response patterns at low levels of τ_{day}). When separated bands of high fitness appear for a location without similarly separated bands of high similarity (e.g. Honolulu), this indicates multiple high-fitness response phenologies.

Figure 8. Genotype x environment interactions mean that different genotypes have greater relative fitness advantages in different years; phenological cueing strategies interact with climatic variation to reduce

consistent fitness advantages. a.) In this example, each color represents an individual genotype from the final generation of a simulation using climatic data from Ithaca, NY, USA, evaluated in a random subset of 20 years. To facilitate interpretation, a random individual is highlighted as a black line. Fitness has been scaled by the maximum observed fitness across all years for this simulation. b.) Across all locations, genotype x environment interactions account for 29% of fitness variation on average. The blue probability density function represents the distribution of mean fitness variance components due to genotype x environment interactions in each of 78 locations. The dashed line shows the mean proportion of variation attributable to genotype x environment interactions.

Figure 9. Under two climate change scenarios, organisms with different environmental histories generally respond earlier, but show variable degrees of advancement and highly variable fitness consequences. The position of each circle represents the mean change in the day of response and the proportional change in geometric mean fitness averaged across all evolved genotypes of 60 simulations for each of 78 locations, relative to the historical climate in that location, represented by a black triangle. The color of each circle represents the historical trait effect of day (T_{day}), indicating the relative use of a climatically invariant cue.

Figure 1

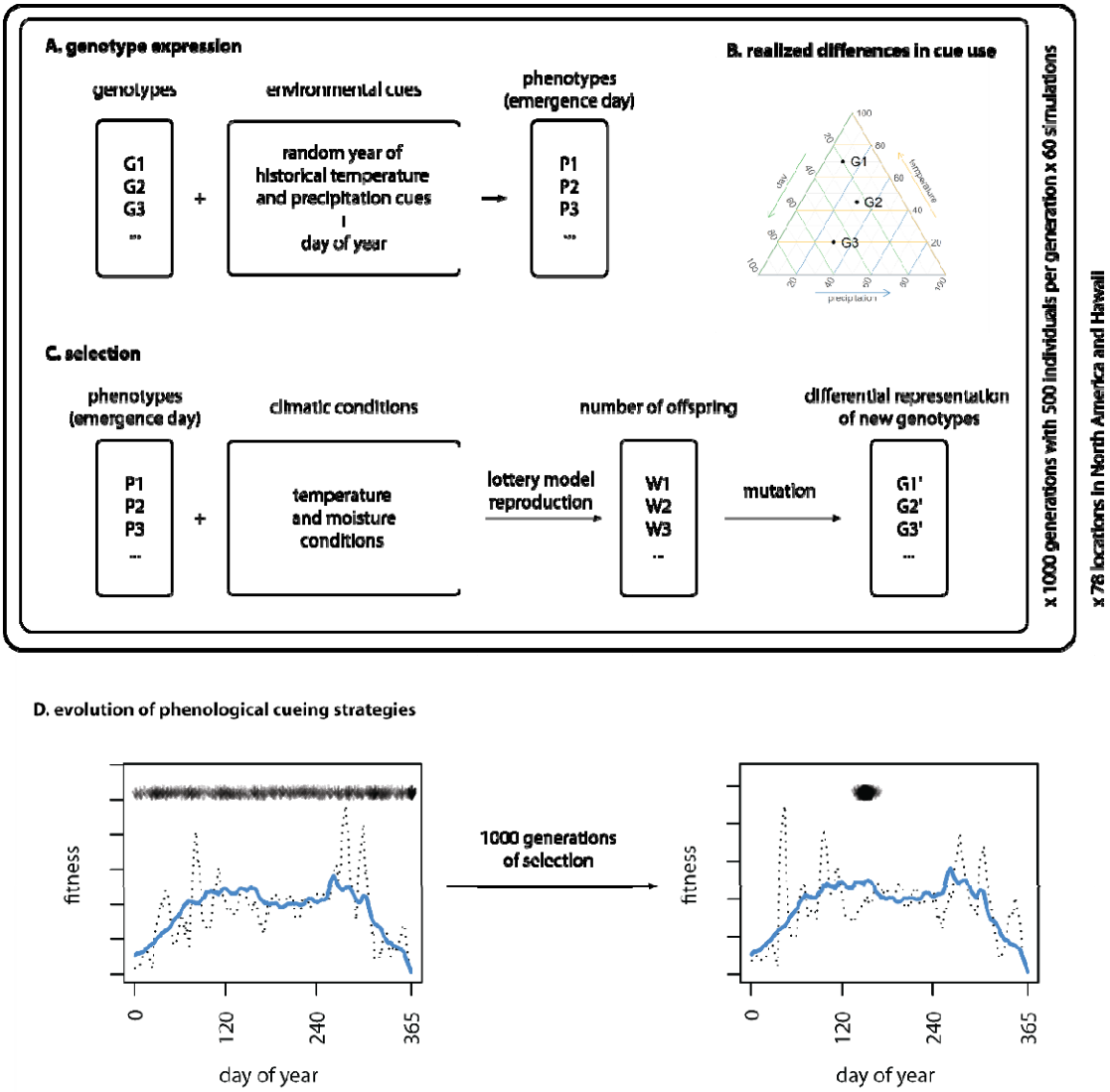

A. genotype expression
genotypes
environmental cues
phenotypes (emergence day)
G1
G2
G3
...
+
random year of historical temperature and precipitation cues
day of year
P1
P2
P3
...
B. realized differences in cue use
G1
G2
G3
C. selection
phenotypes (emergence day)
climatic conditions
number of offspring
differential representation of new genotypes
P1
P2
P3
...
+
temperature and moisture conditions
lottery model reproduction
W1
W2
W3
...
mutation
G1'
G2'
G3'
...
x 1000 generations with 500 individuals per generation x 60 simulations
x 78 locations in North America and Hawaii
D. evolution of phenological cueing strategies
fitness
0
120
240
365
day of year
1000 generations of selection
fitness
0
120
240
365
day of year

Figure 2

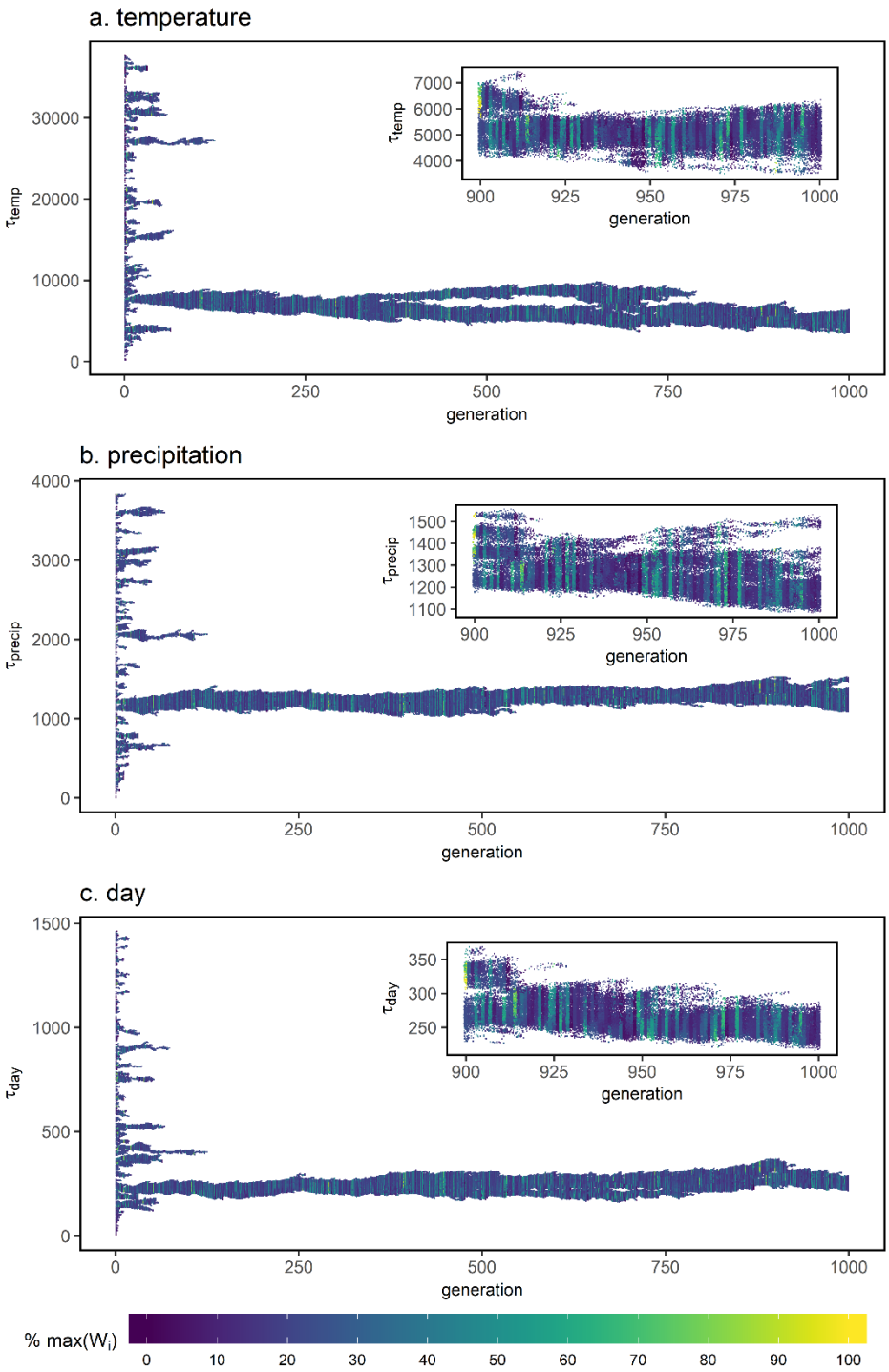

a. temperature
τ_{temp}
generation
b. precipitation
τ_{precip}
generation
c. day
τ_{day}
generation
% max(W_i)

Figure 3

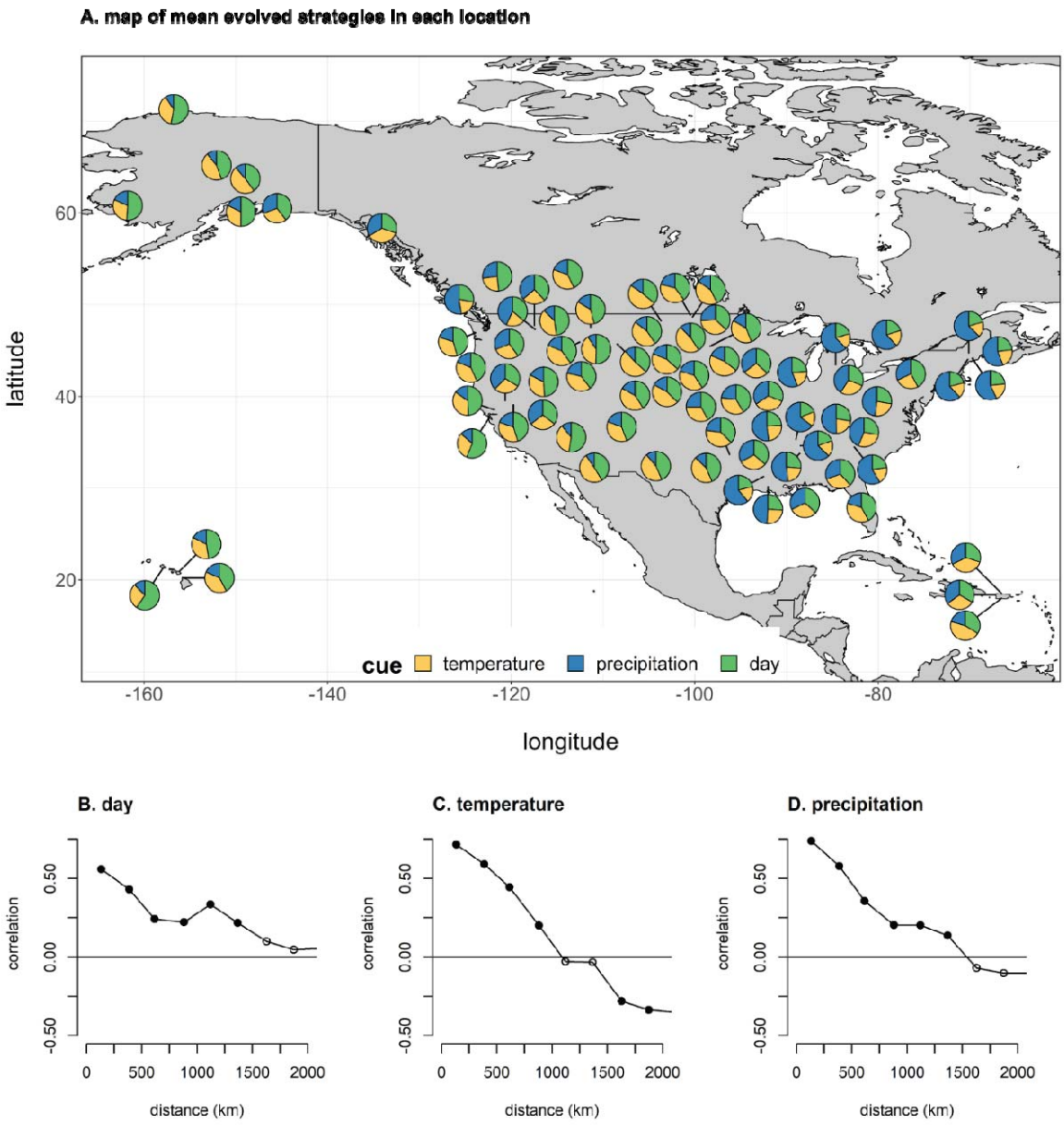

A. map of mean evolved strategies in each location
latitude
longitude
60
40
20
-160
-140
-120
-100
-80
cue
temperature
precipitation
day
B. day
C. temperature
D. precipitation
correlation
0.50
0.00
-0.50
0
500
1000
1500
2000
distance (km)

Figure 4

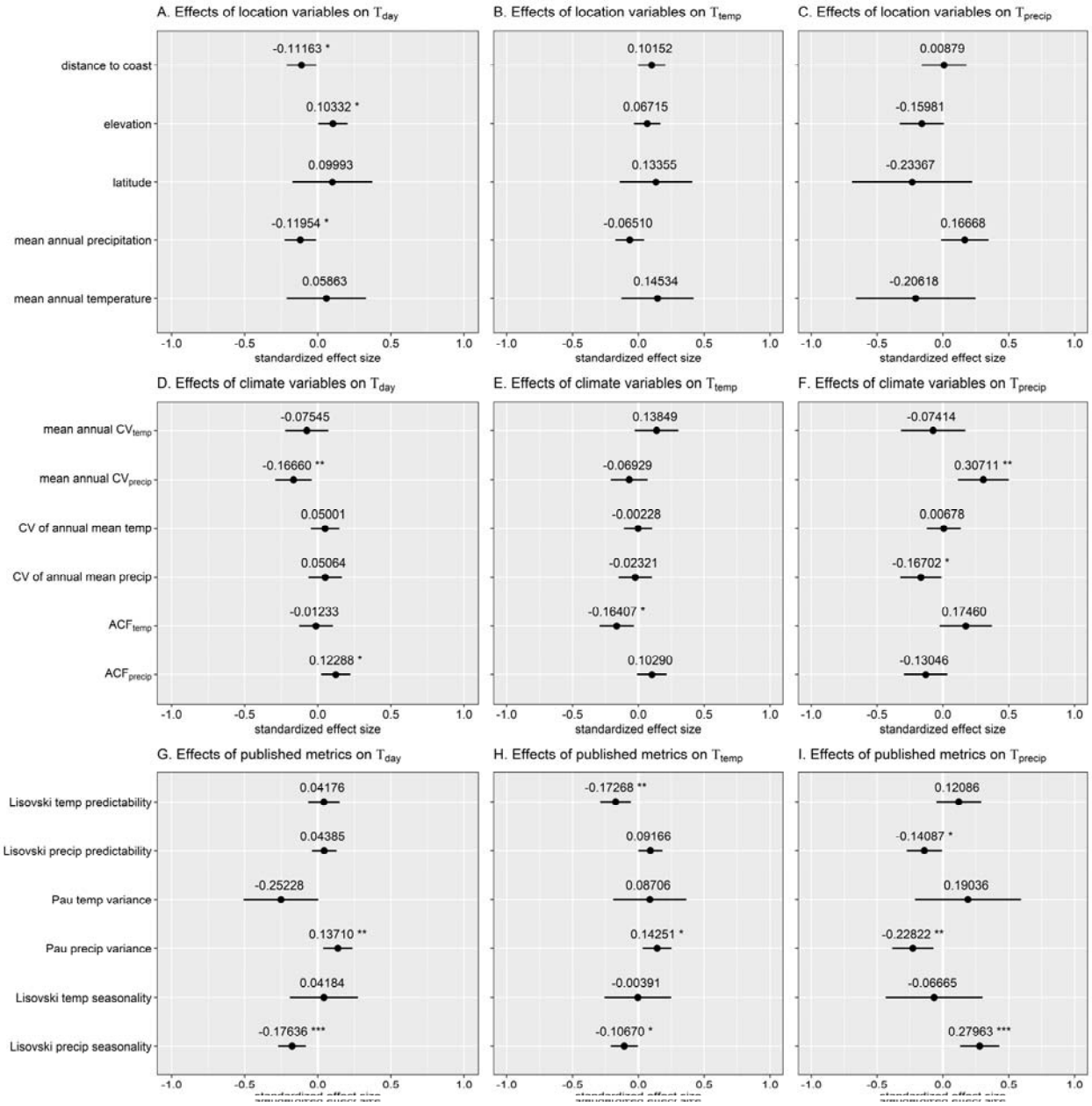

A. Effects of location variables on T_{day}
B. Effects of location variables on T_{temp}
C. Effects of location variables on T_{precip}
distance to coast
elevation
latitude
mean annual precipitation
mean annual temperature
-0.11163 *
0.10332 *
0.09993
-0.11954 *
0.05863
0.10152
0.06715
0.13355
-0.06510
0.14534
0.00879
-0.15981
-0.23367
0.16668
-0.20618
D. Effects of climate variables on T_{day}
E. Effects of climate variables on T_{temp}
F. Effects of climate variables on T_{precip}
mean annual CV_{temp}
mean annual CV_{precip}
CV of annual mean temp
CV of annual mean precip
ACF_{temp}
ACF_{precip}
-0.07545
-0.16660 **
0.05001
0.05064
-0.01233
0.12288 *
0.13849
-0.06929
-0.00228
-0.02321
-0.16407 *
0.10290
-0.07414
0.30711 **
0.00678
-0.16702 *
0.17460
-0.13046
G. Effects of published metrics on T_{day}
H. Effects of published metrics on T_{temp}
I. Effects of published metrics on T_{precip}
Lisovski temp predictability
Lisovski precip predictability
Pau temp variance
Pau precip variance
Lisovski temp seasonality
Lisovski precip seasonality
0.04176
0.04385
-0.25228
0.13710 **
0.04184
-0.17636 ***
-0.17268 **
0.09166
0.08706
0.14251 *
-0.00391
-0.10670 *
0.12086
-0.14087 *
0.19036
-0.22822 **
-0.06665
0.27963 ***
-1.0
-0.5
0.0
0.5
1.0
standardized effect size

Figure 5

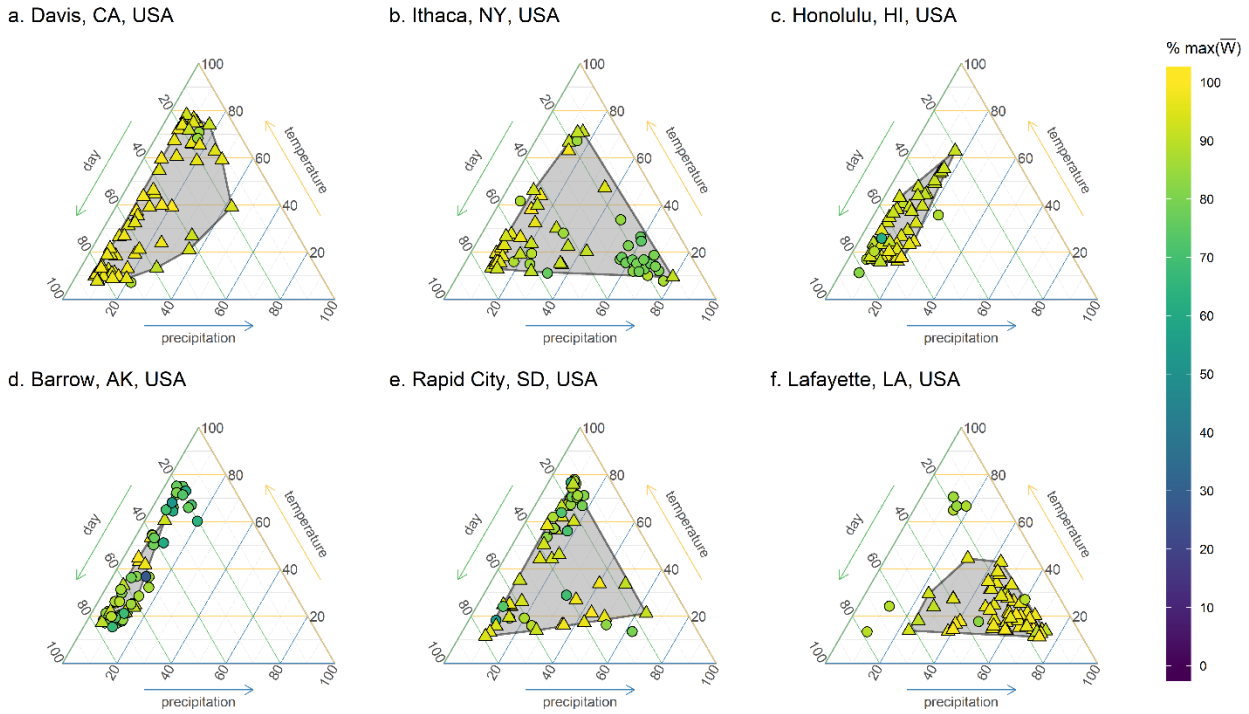

a. Davis, CA, USA
b. Ithaca, NY, USA
c. Honolulu, HI, USA
d. Barrow, AK, USA
e. Rapid City, SD, USA
f. Lafayette, LA, USA
day
temperature
precipitation
% max($\overline{W}$)

Figure 6

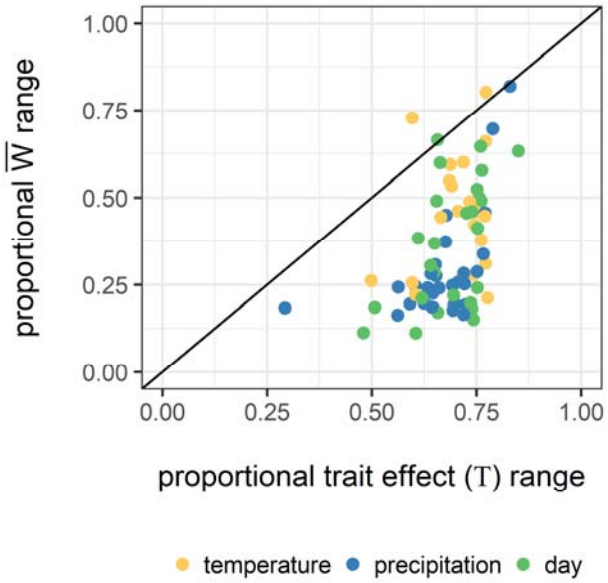
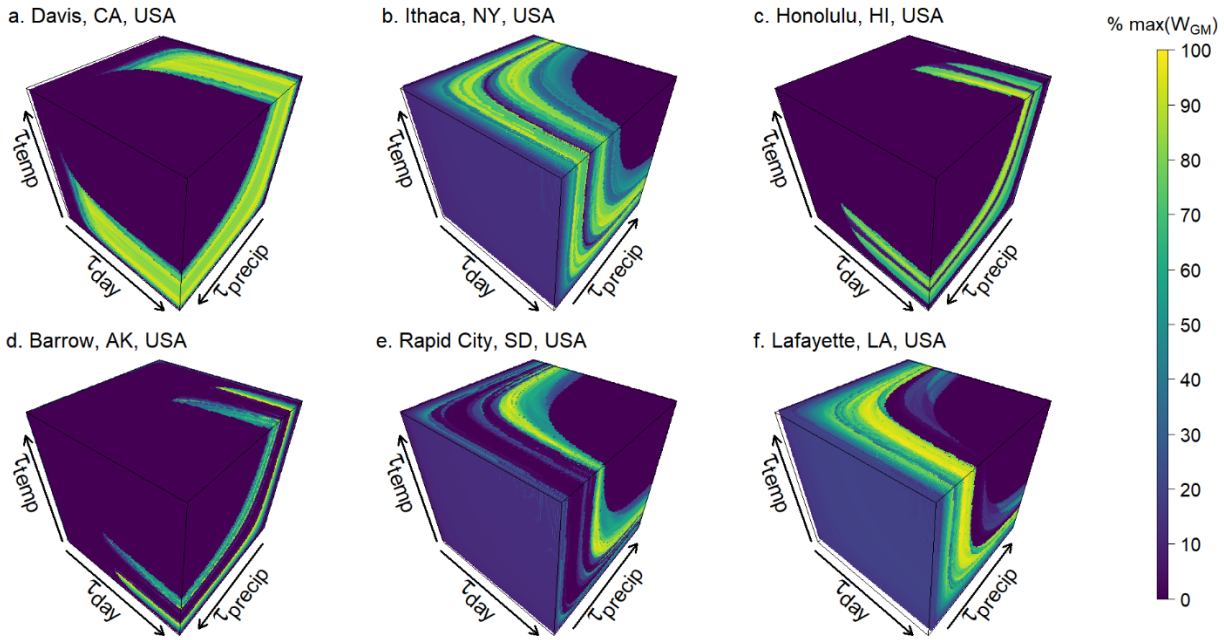

1.00
0.75
0.50
0.25
0.00
proportional $\overline{W}$ range
0.00
0.25
0.50
0.75
1.00
proportional trait effect (T) range
temperature
precipitation
day

Figure 7

A)


a. Davis, CA, USA
b. Ithaca, NY, USA
c. Honolulu, HI, USA
d. Barrow, AK, USA
e. Rapid City, SD, USA
f. Lafayette, LA, USA
τtemp
τday
τprecip
% max(WGM)
100
90
80
70
60
50
40
30
20
10
0

B)

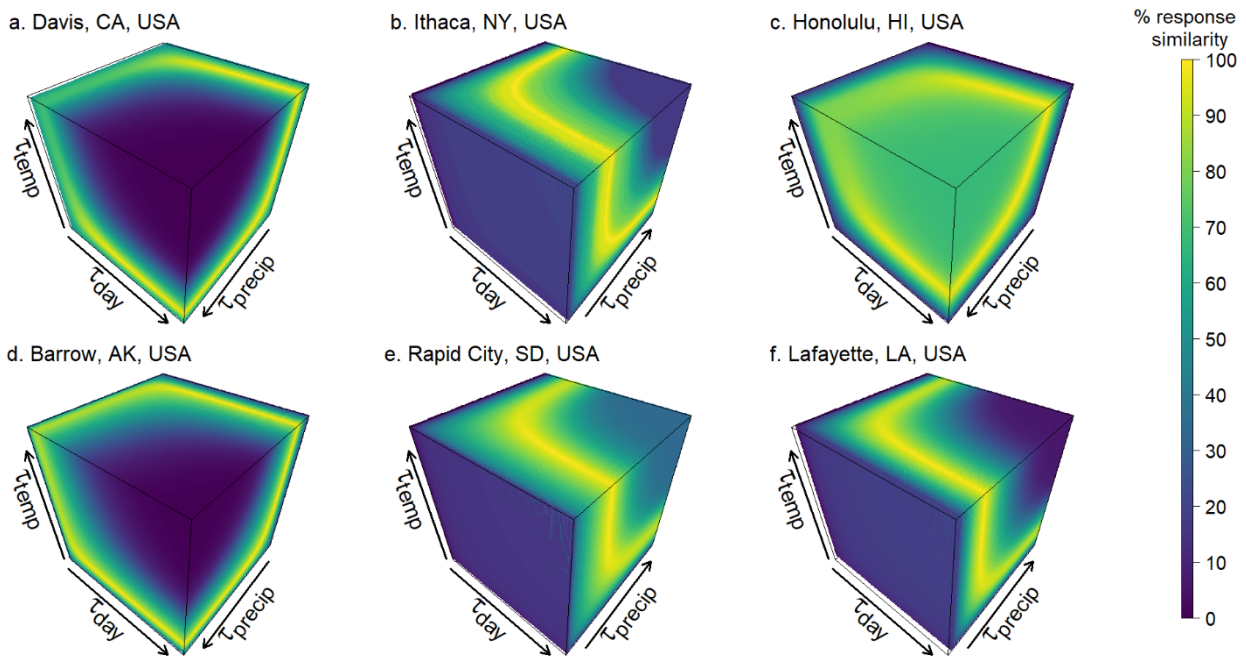

a. Davis, CA, USA
b. Ithaca, NY, USA
c. Honolulu, HI, USA
d. Barrow, AK, USA
e. Rapid City, SD, USA
f. Lafayette, LA, USA
τtemp
τday
τprecip
% response similarity
100
90
80
70
60
50
40
30
20
10
0

Figure 8

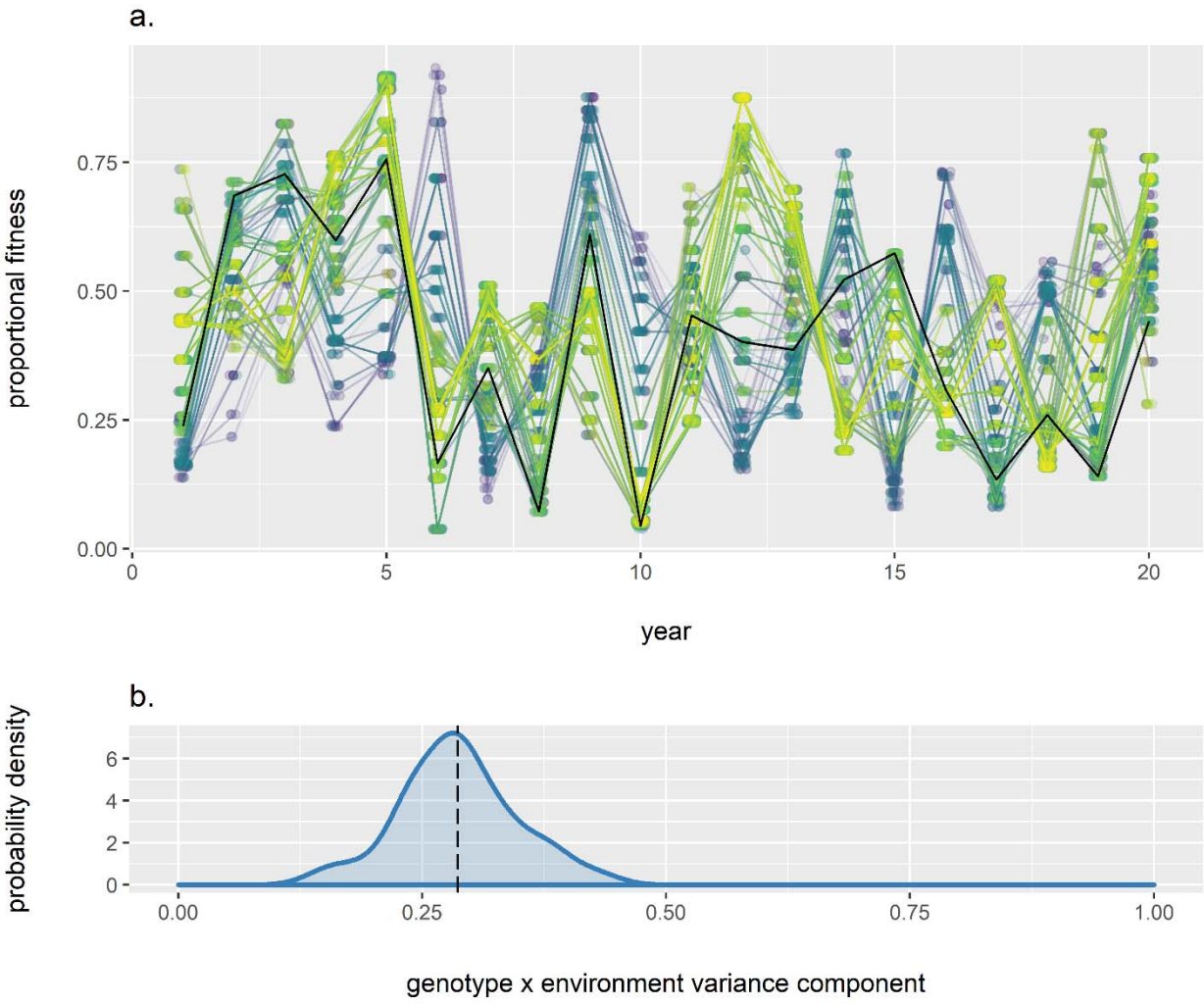
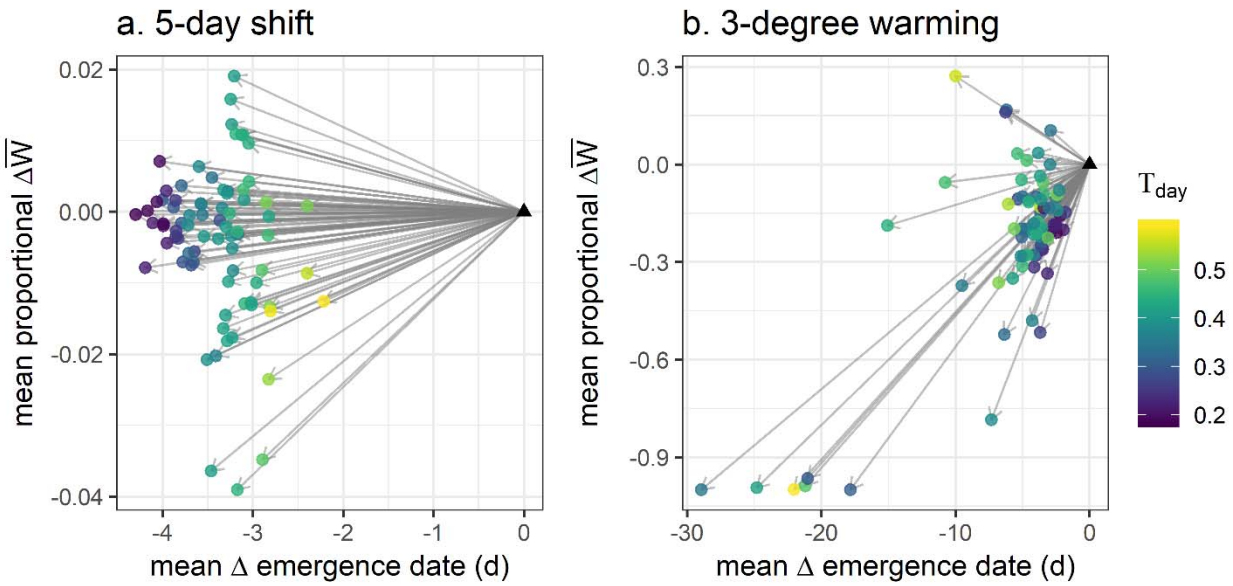

a.
proportional fitness
0.75
0.50
0.25
0.00
0
5
10
15
20
year
b.
probability density
6
4
2
0
0.00
0.25
0.50
0.75
1.00
genotype x environment variance component

Figure 9


a. 5-day shift
b. 3-degree warming
mean proportional ΔW̄
mean Δ emergence date (d)
0.02
0.00
-0.02
-0.04
-4
-3
-2
-1
0
0.3
0.0
-0.3
-0.6
-0.9
-30
-20
-10
0
T day
0.5
0.4
0.3
0.2

Supporting Data and Material

Supporting Figures S1-S6

Supporting Methods.

Supporting Table 1. Location data

Supporting figure legends

Figure S1. A map of mean phenological cueing strategies using trait effects in different locations under an alternative model formulation. To generate this alternative map, our model used cumulative photoperiod as the cue for day of year, and we imposed a developmental baseline temperature of 0 C across all locations below which organisms are insensitive to thermal cues. As in the primary model formulation, similar strategies evolve in similar climates, which can be seen as spatial autocorrelation in this map.

Figure S2. Evolved strategies for 60 simulations in each of 78 locations. Each point represents the mean strategy at the end of one simulation; each strategy is represented as a composition of the "trait effects" in percentages, which represent relative cue use (see "assessing realized relative cue use" in Methods). Point color reflects the mean geometric mean fitness for each simulation scaled by location. Simulations within 10% of the maximum observed geometric mean fitness in each location are represented as triangles and included in a gray convex hull. All other points are represented as circles.

Figure S3. Simple hypothesized explanatory variables were not strongly correlated with evolved strategies. Values indicate raw, unstandardized fixed effect coefficient estimates from linear models (see "Analysis of explanatory factors"), error bars indicate 95% confidence limits.

Figure S4. Strategy variation across locations. Strategy variation in each location is measured as the compositional metric standard deviation A) among all genotypes and B) among all genotypes within 10% of the most fit per location.

Figure S5. Most simulations result in evolved strategies that have geometric mean fitnesses that are within 10% of the most fit simulation in each location.

Figure S6. Location specific responses to A) a 5-day shift in temperature and precipitation regimes and B) 3-degree C warming. For each location, each of 30 genotypes is represented by a blue and red circle showing its day of response and proportional fitness under historical and changed conditions, respectively.

Figure S1

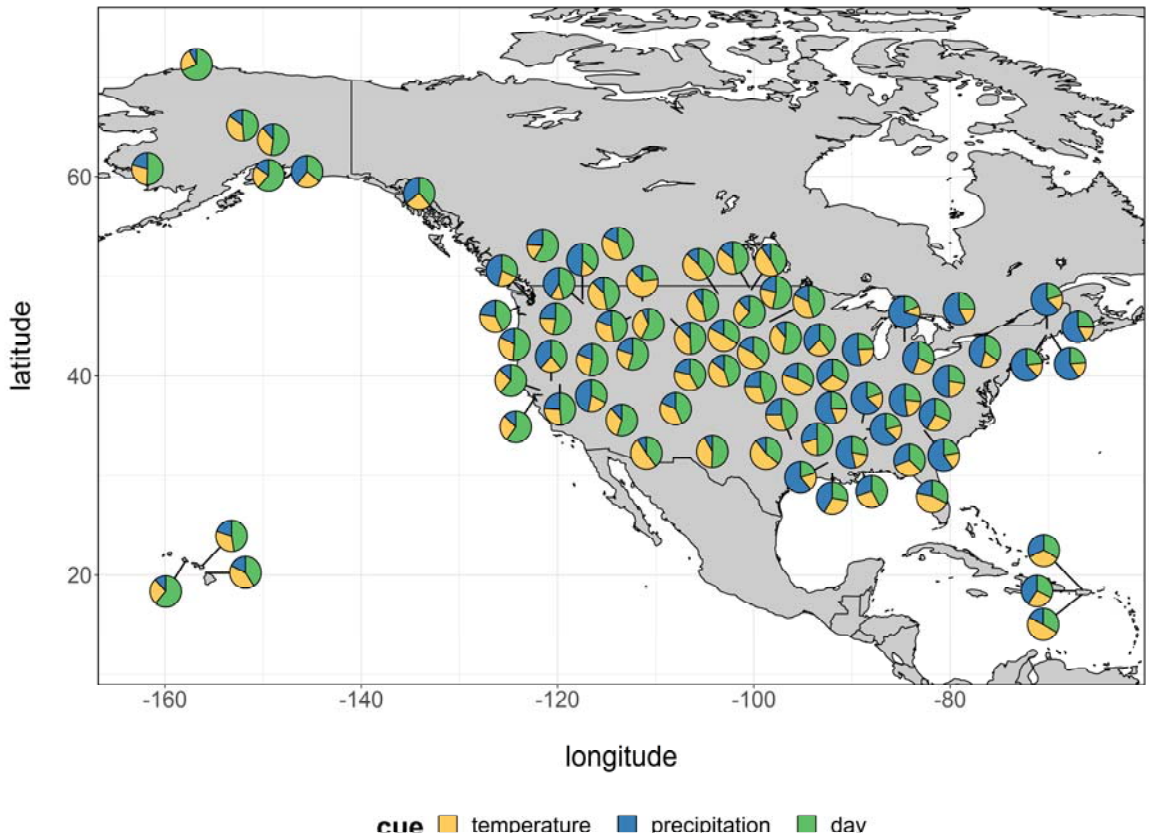
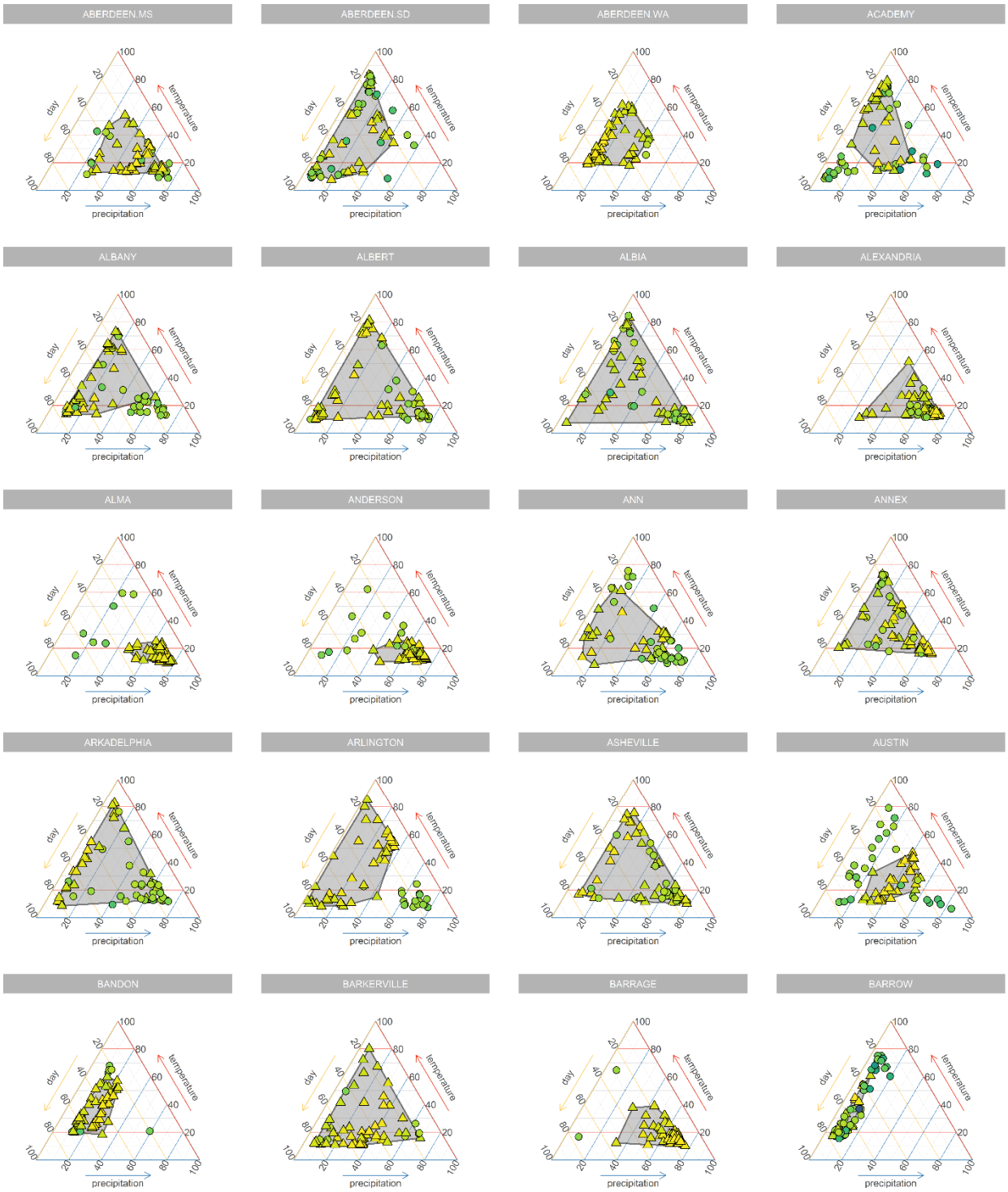
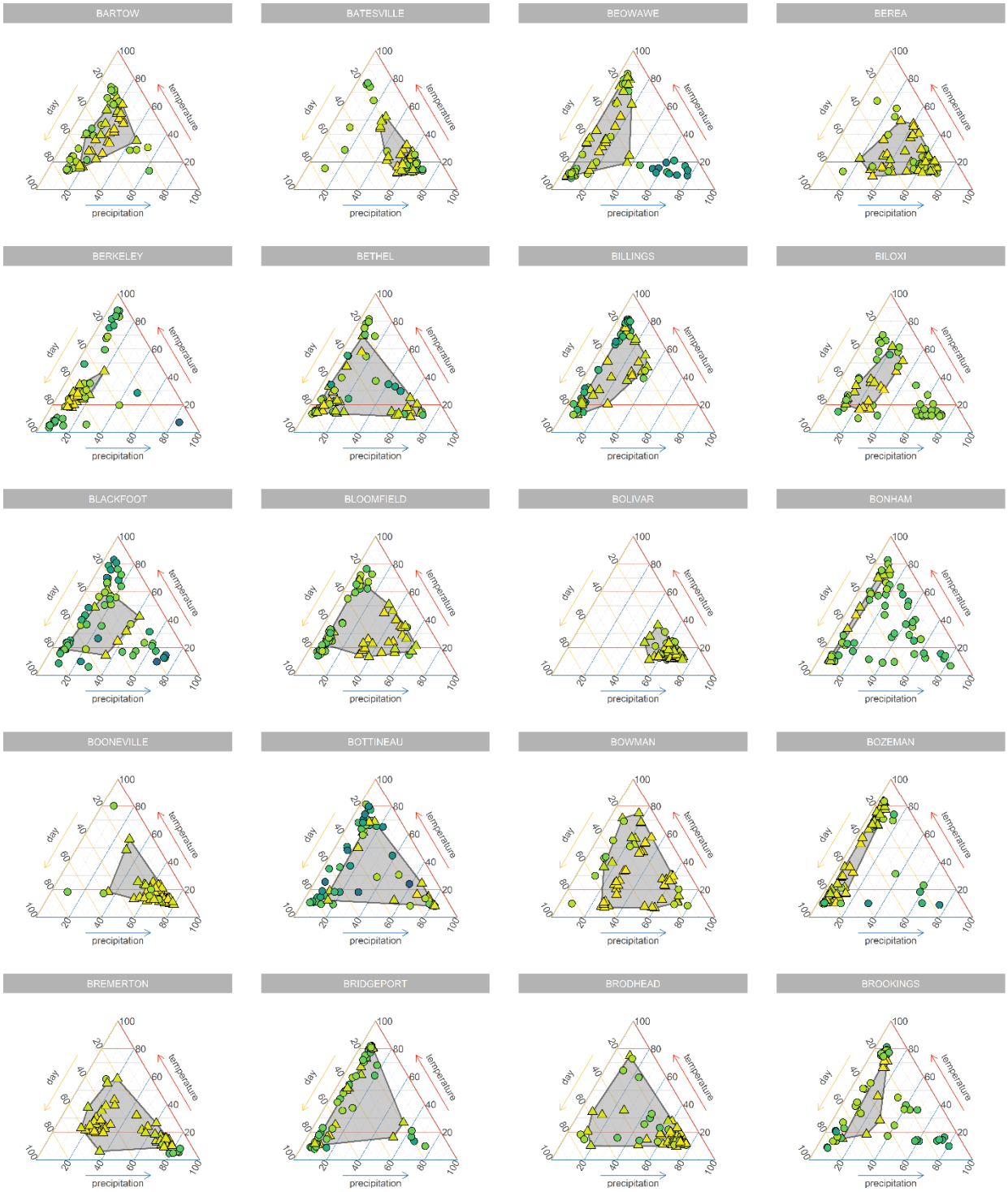

60
40
20
latitude
-160
-140
-120
-100
-80
longitude
cue
temperature
precipitation
day

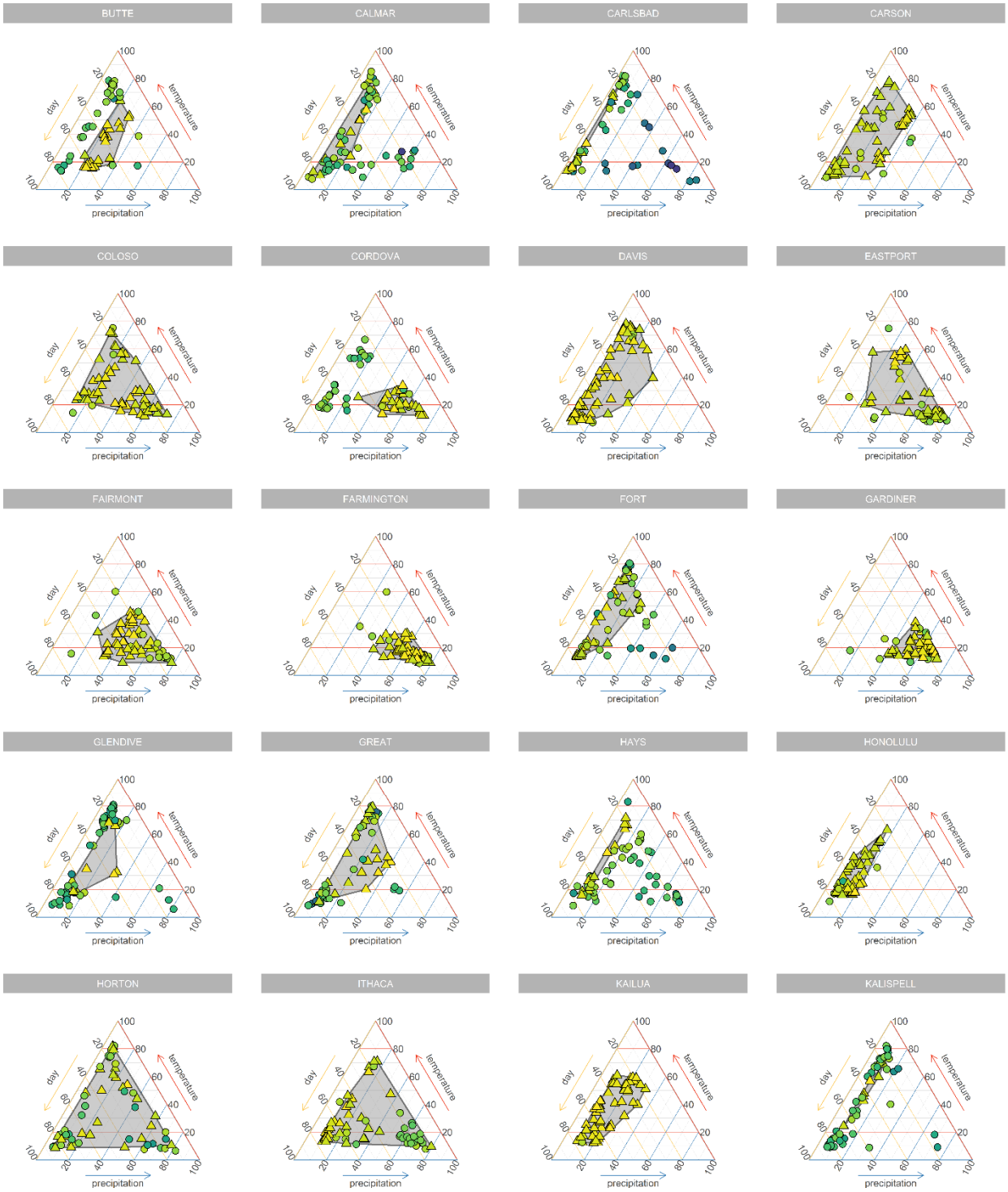
Figure S2


ABERDEEN.MS
ABERDEEN.SD
ABERDEEN.WA
ACADEMY
ALBANY
ALBERT
ALBIA
ALEXANDRIA
ALMA
ANDERSON
ANN
ANNEX
ARKADELPHIA
ARLINGTON
ASHEVILLE
AUSTIN
BANDON
BARKERVILLE
BARRAGE
BARROW
day
temperature
precipitation


BARTOW
BATESVILLE
BEOWAWE
BEREA
BERKELEY
BETHEL
BILLINGS
BILOXI
BLACKFOOT
BLOOMFIELD
BOLIVAR
BONHAM
BOONEVILLE
BOTTINEAU
BOWMAN
BOZEMAN
BREMERTON
BRIDGEPORT
BRODHEAD
BROOKINGS
day
temperature
precipitation

848

849


BUTTE
CALMAR
CARLSBAD
CARSON
COLOSO
CORDOVA
DAVIS
EASTPORT
FAIRMONT
FARMINGTON
FORT
GARDINER
GLENDIVE
GREAT
HAYS
HONOLULU
HORTON
ITHACA
KAILUA
KALISPELL
day
temperature
precipitation

850

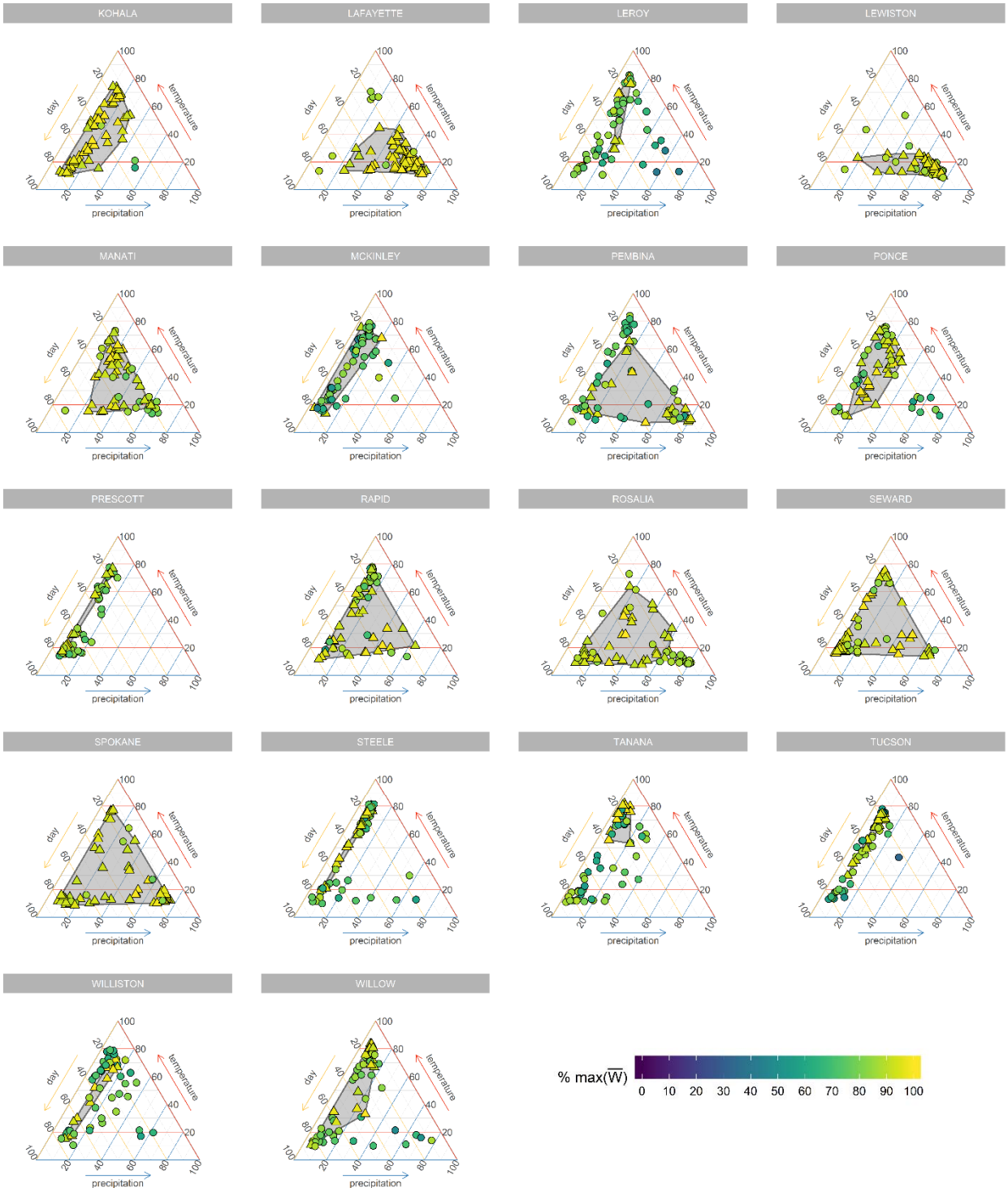

KOHALA
LAFAYETTE
LEROY
LEWISTON
MANATI
MCKINLEY
PEMBINA
PONCE
PRESCOTT
RAPID
ROSALIA
SEWARD
SPOKANE
STEELE
TANANA
TUCSON
WILLISTON
WILLOW
day
temperature
precipitation
% max($\overline{W}$)
0 10 20 30 40 50 60 70 80 90 100

Figure S3

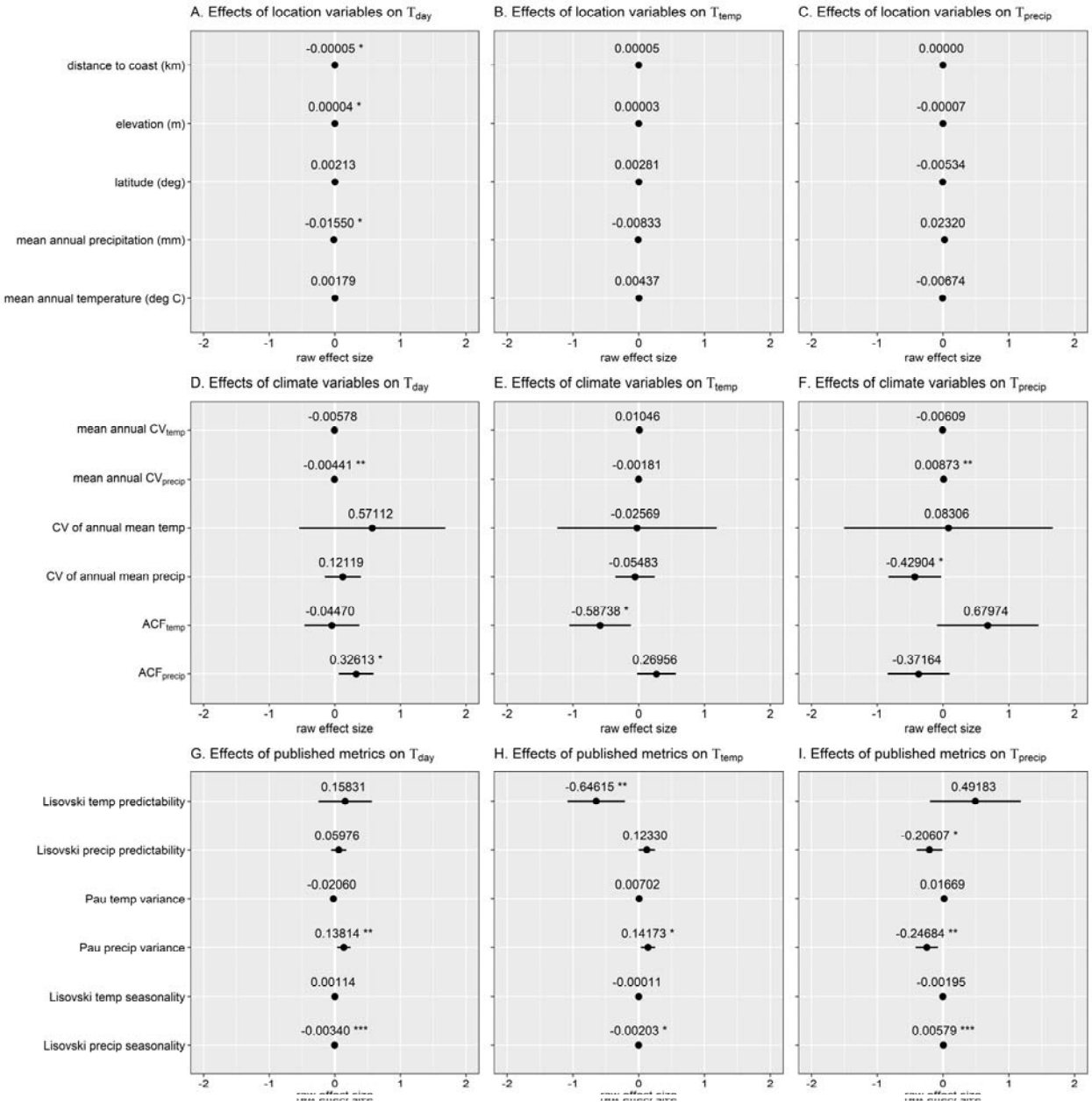

A. Effects of location variables on T_{day}
B. Effects of location variables on T_{temp}
C. Effects of location variables on T_{precip}
distance to coast (km)
elevation (m)
latitude (deg)
mean annual precipitation (mm)
mean annual temperature (deg C)
-0.00005 *
0.00004 *
0.00213
-0.01550 *
0.00179
0.00005
0.00003
0.00281
-0.00833
0.00437
0.00000
-0.00007
-0.00534
0.02320
-0.00674
raw effect size
D. Effects of climate variables on T_{day}
E. Effects of climate variables on T_{temp}
F. Effects of climate variables on T_{precip}
mean annual CV_{temp}
mean annual CV_{precip}
CV of annual mean temp
CV of annual mean precip
ACF_{temp}
ACF_{precip}
-0.00578
-0.00441 **
0.57112
0.12119
-0.04470
0.32613 *
0.01046
-0.00181
-0.02569
-0.05483
-0.58738 *
0.26956
-0.00609
0.00873 **
0.08306
-0.42904 *
0.67974
-0.37164
G. Effects of published metrics on T_{day}
H. Effects of published metrics on T_{temp}
I. Effects of published metrics on T_{precip}
Lisovski temp predictability
Lisovski precip predictability
Pau temp variance
Pau precip variance
Lisovski temp seasonality
Lisovski precip seasonality
0.15831
0.05976
-0.02060
0.13814 **
0.00114
-0.00340 ***
-0.64615 **
0.12330
0.00702
0.14173 *
-0.00011
-0.00203 *
0.49183
-0.20607 *
0.01669
-0.24684 **
-0.00195
0.00579 ***

Figure S4

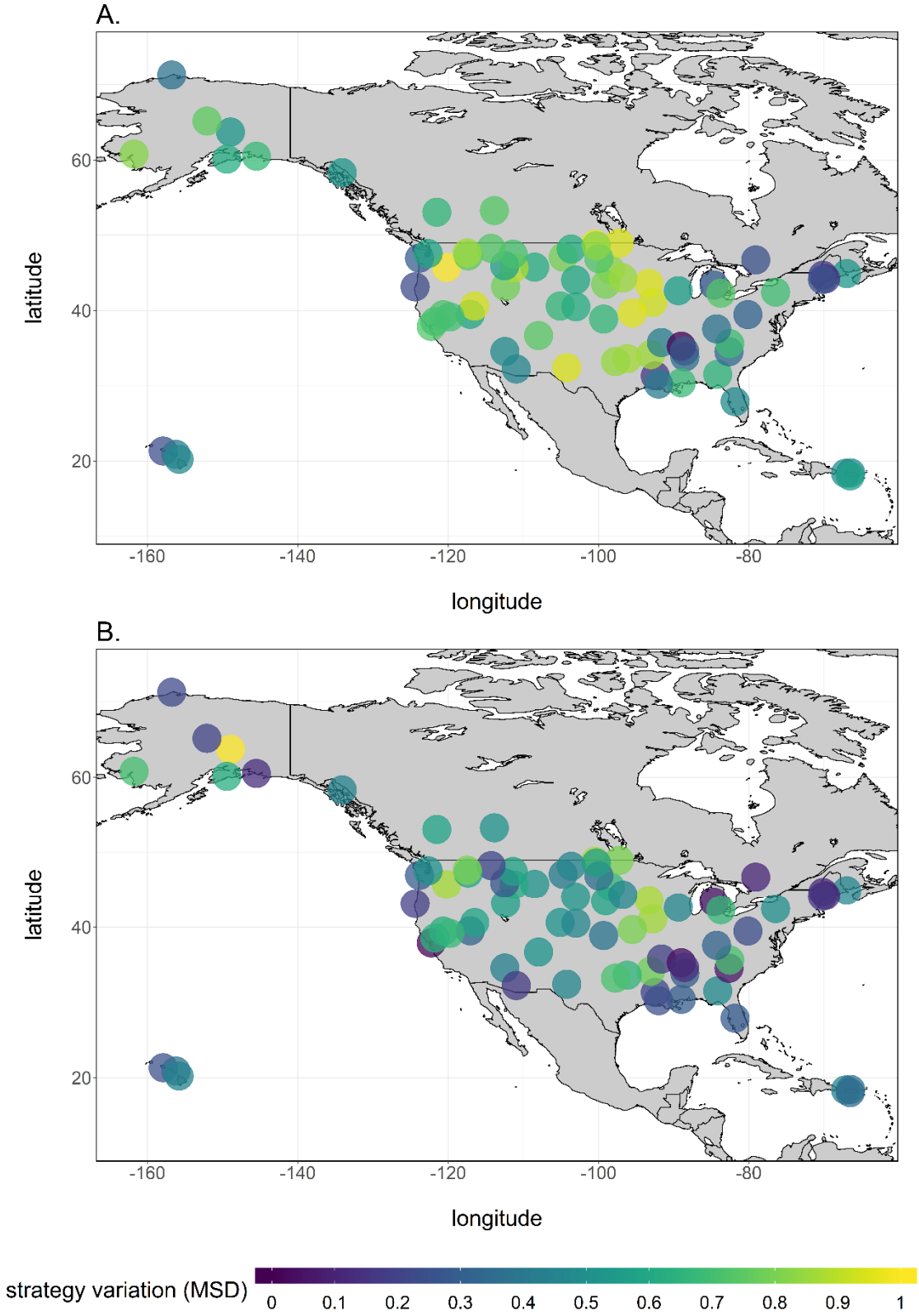
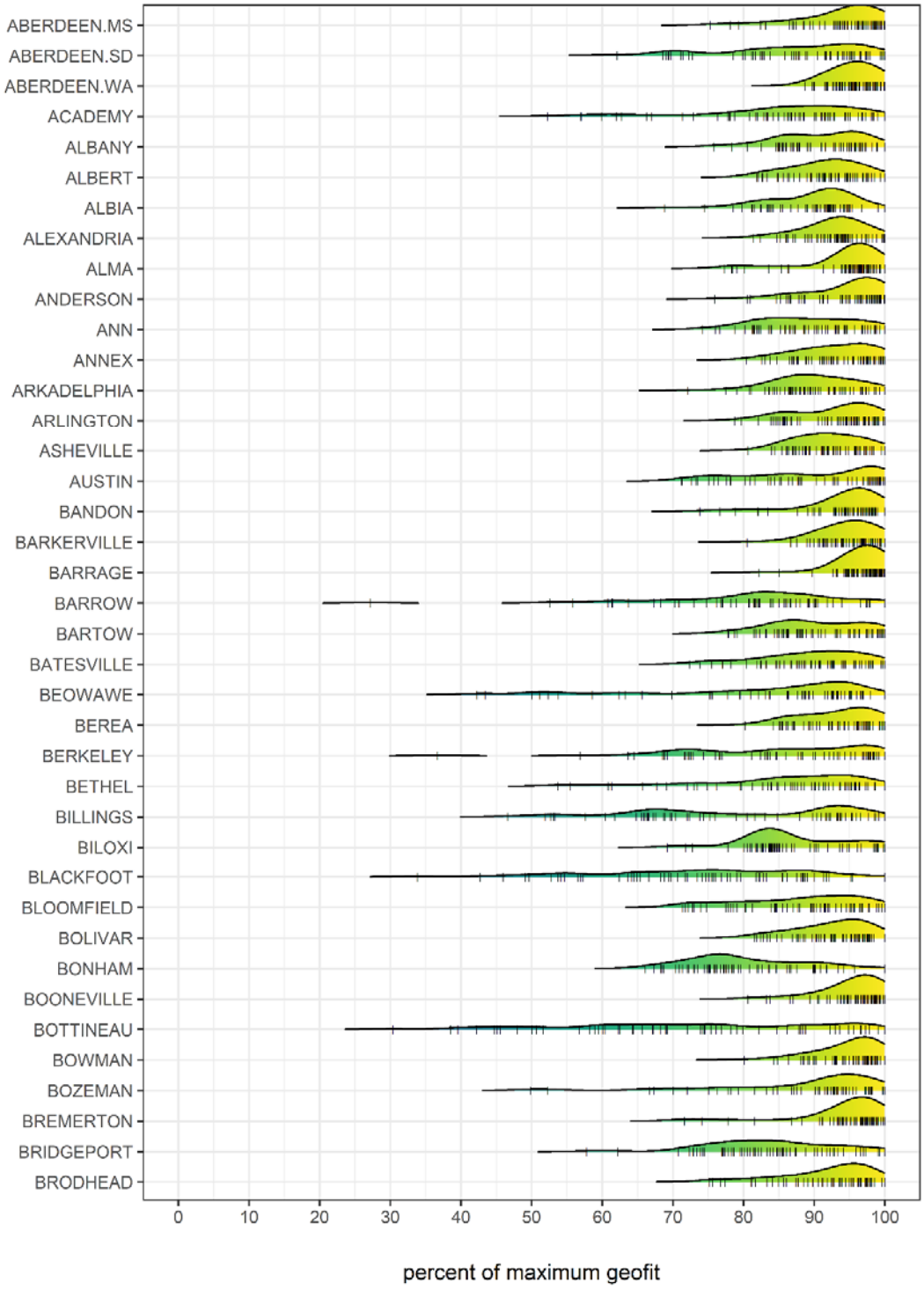

A.
latitude
60
40
20
-160
-140
-120
-100
-80
longitude
B.
latitude
60
40
20
-160
-140
-120
-100
-80
longitude
strategy variation (MSD)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure S5


ABERDEEN.MS
ABERDEEN.SD
ABERDEEN.WA
ACADEMY
ALBANY
ALBERT
ALBIA
ALEXANDRIA
ALMA
ANDERSON
ANN
ANNEX
ARKADELPHIA
ARLINGTON
ASHEVILLE
AUSTIN
BANDON
BARKERVILLE
BARRAGE
BARROW
BARTOW
BATESVILLE
BEOWAWE
BEREA
BERKELEY
BETHEL
BILLINGS
BILOXI
BLACKFOOT
BLOOMFIELD
BOLIVAR
BONHAM
BOONEVILLE
BOTTINEAU
BOWMAN
BOZEMAN
BREMERTON
BRIDGEPORT
BRODHEAD
0
10
20
30
40
50
60
70
80
90
100
percent of maximum geofit

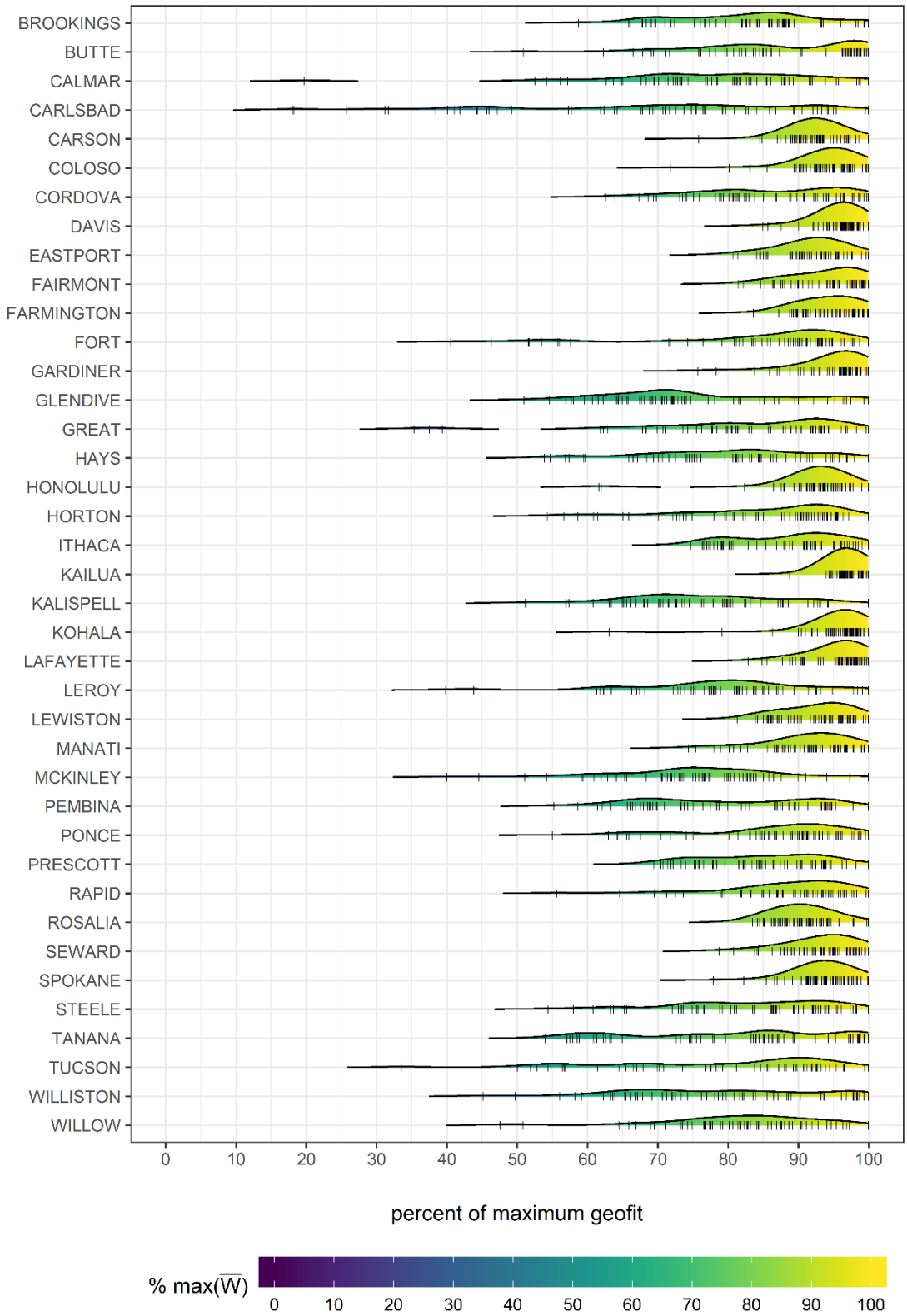
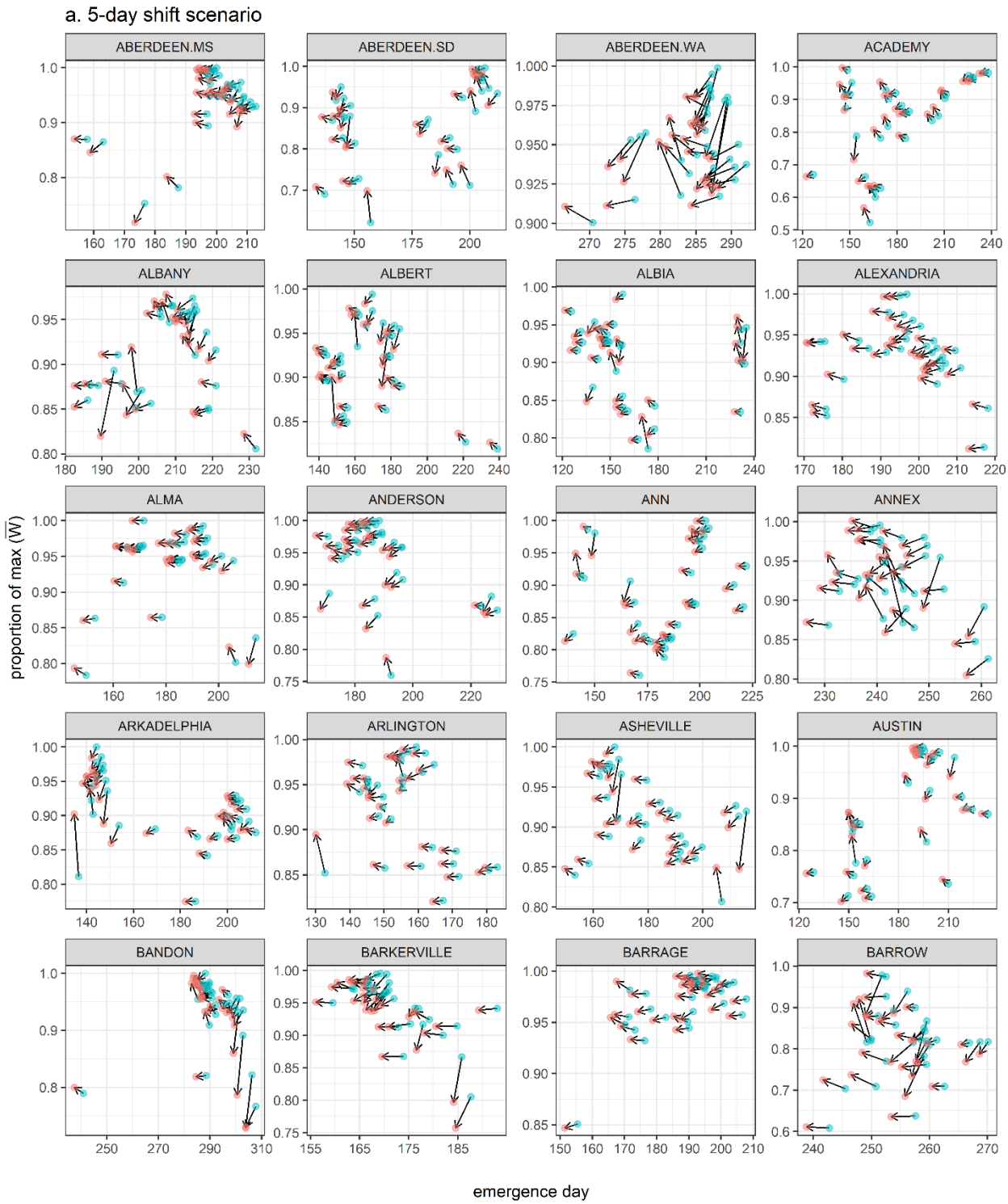
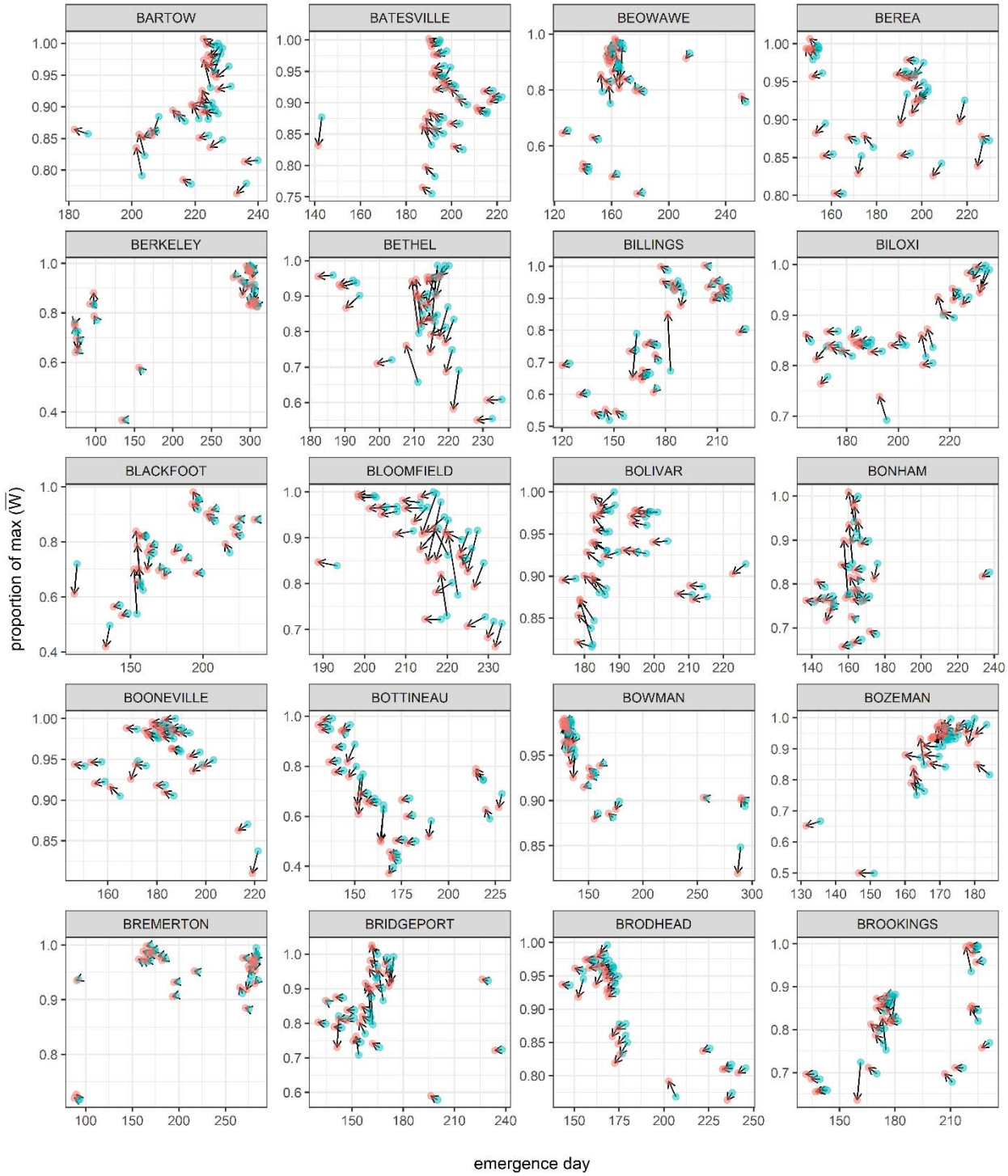
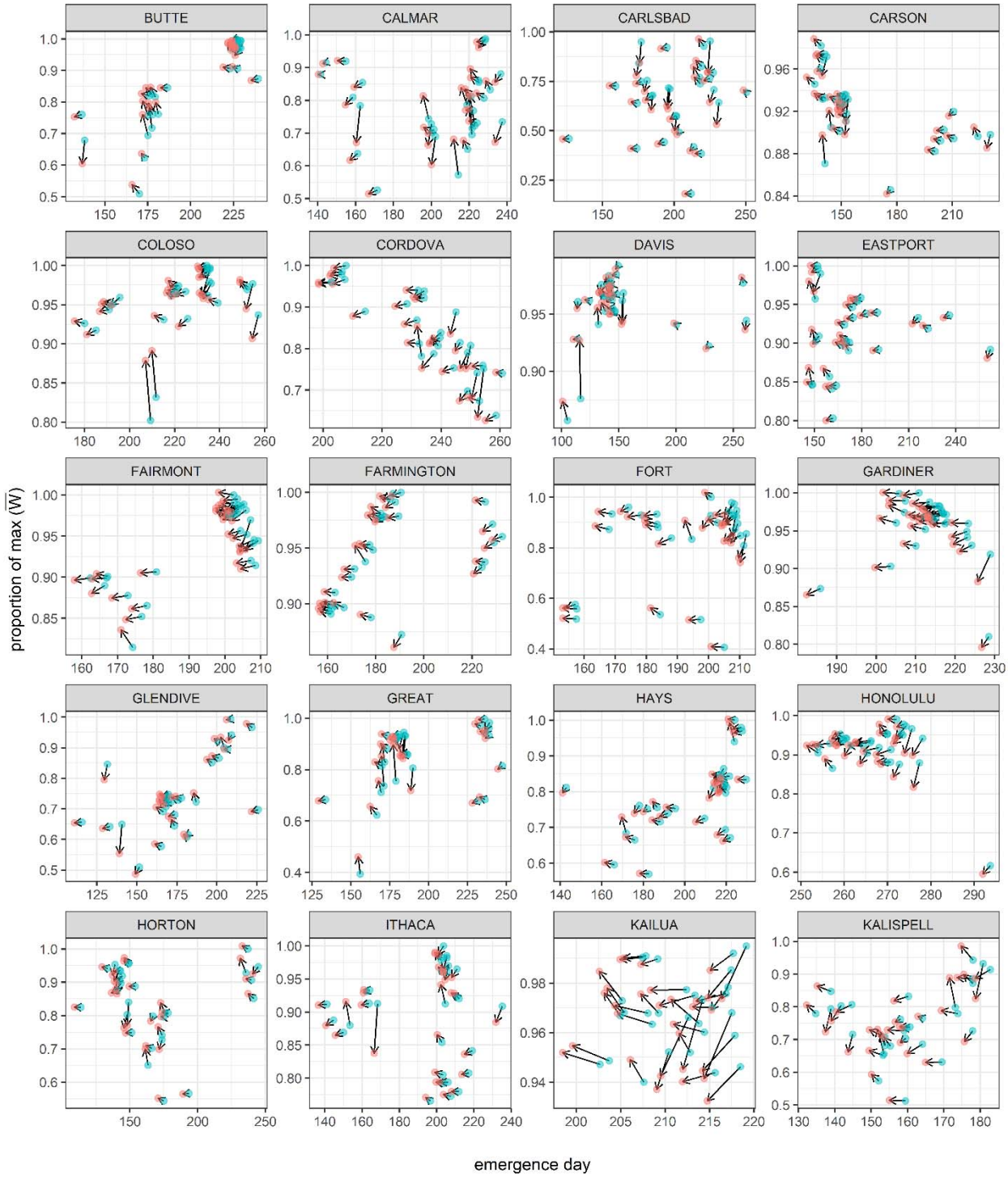

BROOKINGS
BUTTE
CALMAR
CARLSBAD
CARSON
COLOSO
CORDOVA
DAVIS
EASTPORT
FAIRMONT
FARMINGTON
FORT
GARDINER
GLENDIVE
GREAT
HAYS
HONOLULU
HORTON
ITHACA
KAILUA
KALISPELL
KOHALA
LAFAYETTE
LEROY
LEWISTON
MANATI
MCKINLEY
PEMBINA
PONCE
PRESCOTT
RAPID
ROSALIA
SEWARD
SPOKANE
STEELE
TANANA
TUCSON
WILLISTON
WILLOW
0 10 20 30 40 50 60 70 80 90 100
percent of maximum geofit
% max($\overline{W}$)
0 10 20 30 40 50 60 70 80 90 100

Figure S6

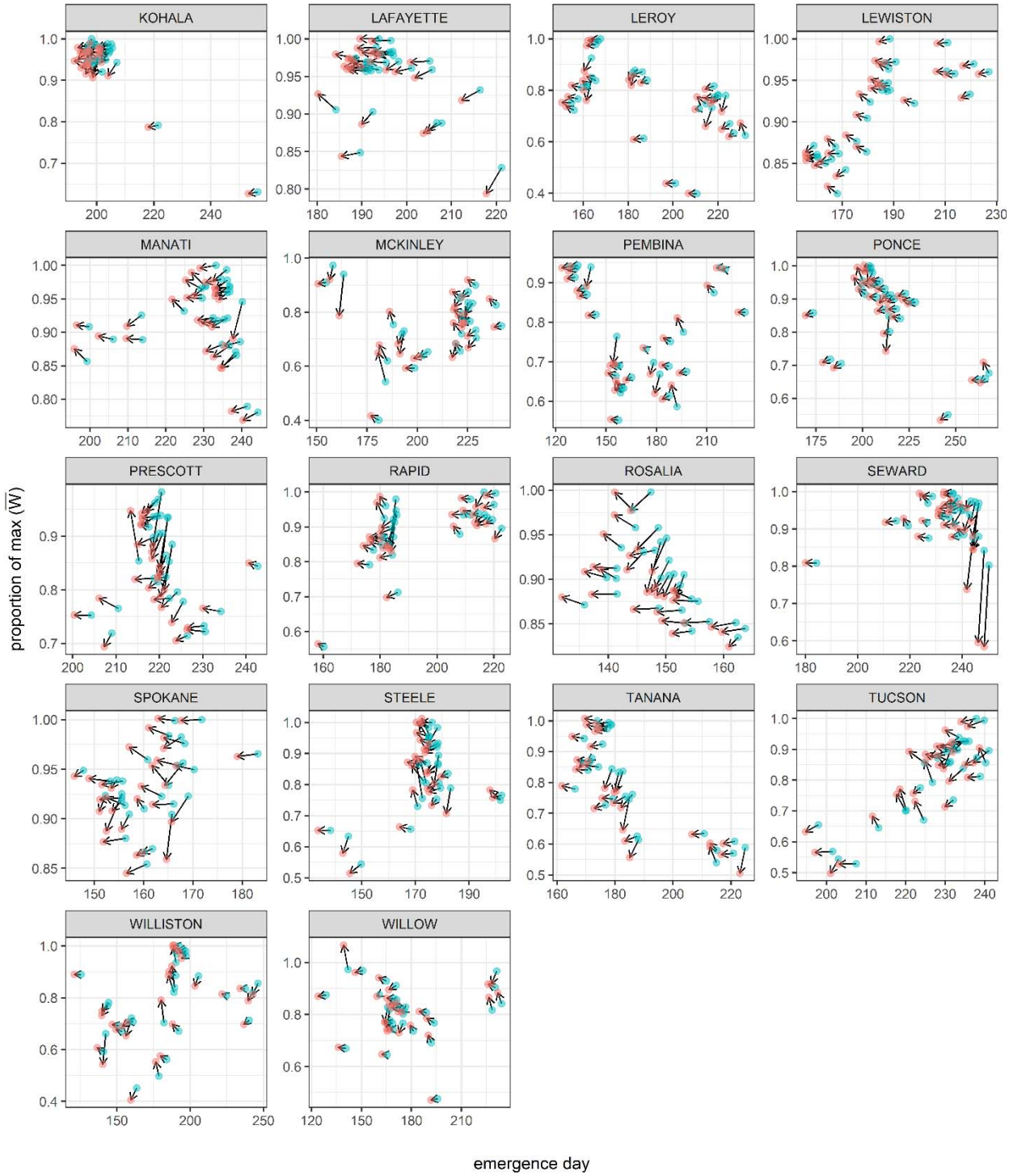

a. 5-day shift scenario
ABERDEEN.MS
ABERDEEN.SD
ABERDEEN.WA
ACADEMY
ALBANY
ALBERT
ALBIA
ALEXANDRIA
ALMA
ANDERSON
ANN
ANNEX
ARKADELPHIA
ARLINGTON
ASHEVILLE
AUSTIN
BANDON
BARKERVILLE
BARRAGE
BARROW
proportion of max ($\overline{W}$)
emergence day


BARTOW
BATESVILLE
BEOWAWE
BEREA
BERKELEY
BETHEL
BILLINGS
BILOXI
BLACKFOOT
BLOOMFIELD
BOLIVAR
BONHAM
BOONEVILLE
BOTTINEAU
BOWMAN
BOZEMAN
BREMERTON
BRIDGEPORT
BRODHEAD
BROOKINGS
proportion of max (W̄)
emergence day

866


BUTTE
CALMAR
CARLSBAD
CARSON
COLOSO
CORDOVA
DAVIS
EASTPORT
FAIRMONT
FARMINGTON
FORT
GARDINER
GLENDIVE
GREAT
HAYS
HONOLULU
HORTON
ITHACA
KAILUA
KALISPELL
proportion of max (W)
emergence day

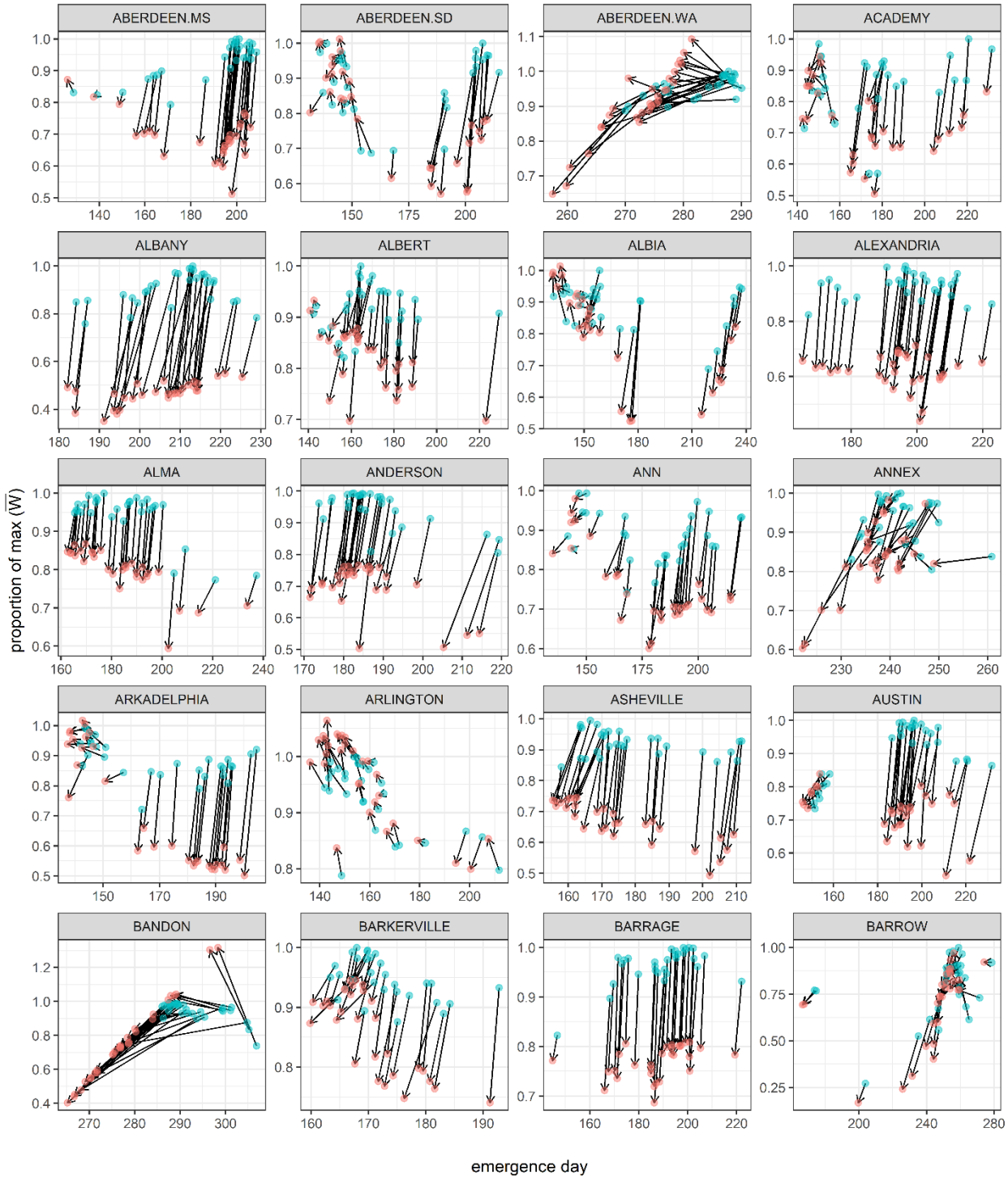
867


KOHALA
LAFAYETTE
LEROY
LEWISTON
MANATI
MCKINLEY
PEMBINA
PONCE
PRESCOTT
RAPID
ROSALIA
SEWARD
SPOKANE
STEELE
TANANA
TUCSON
WILLISTON
WILLOW
proportion of max (W)
emergence day

868

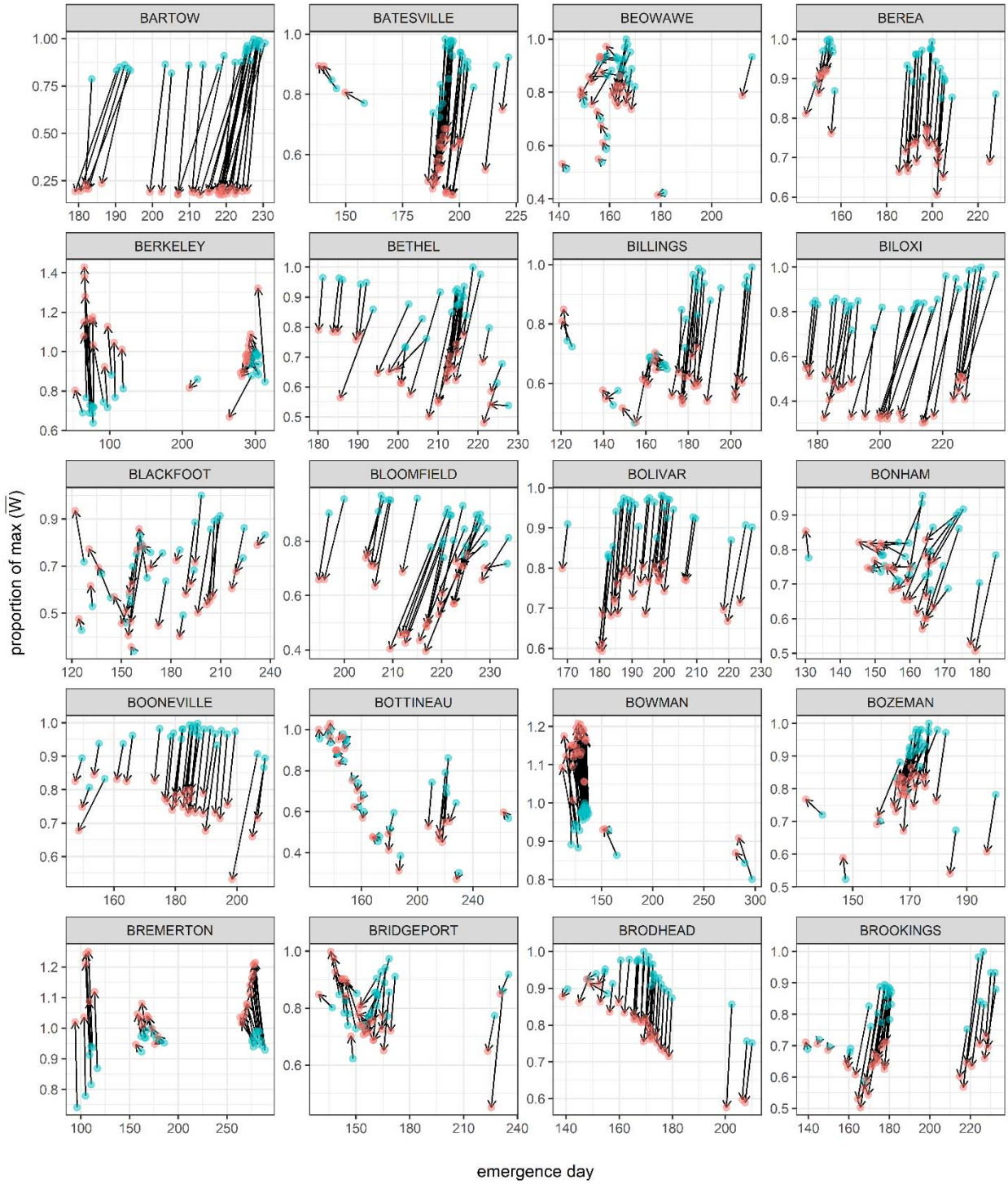
869

b. 3-degree warming scenario



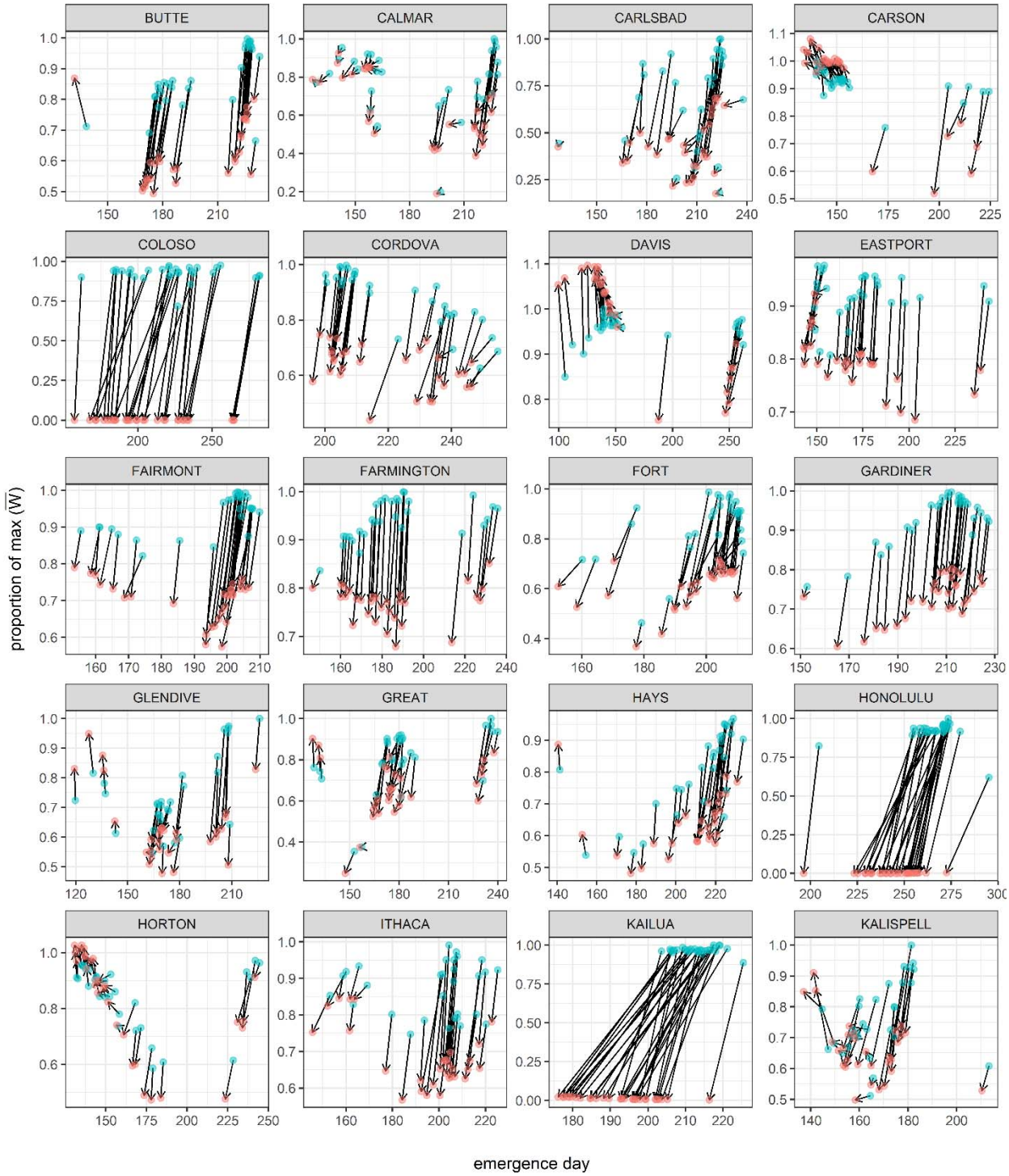
870

871


BARTOW
BATESVILLE
BEOWAWE
BEREA
BERKELEY
BETHEL
BILLINGS
BILOXI
BLACKFOOT
BLOOMFIELD
BOLIVAR
BONHAM
BOONEVILLE
BOTTINEAU
BOWMAN
BOZEMAN
BREMERTON
BRIDGEPORT
BRODHEAD
BROOKINGS
proportion of max (W̄)
emergence day

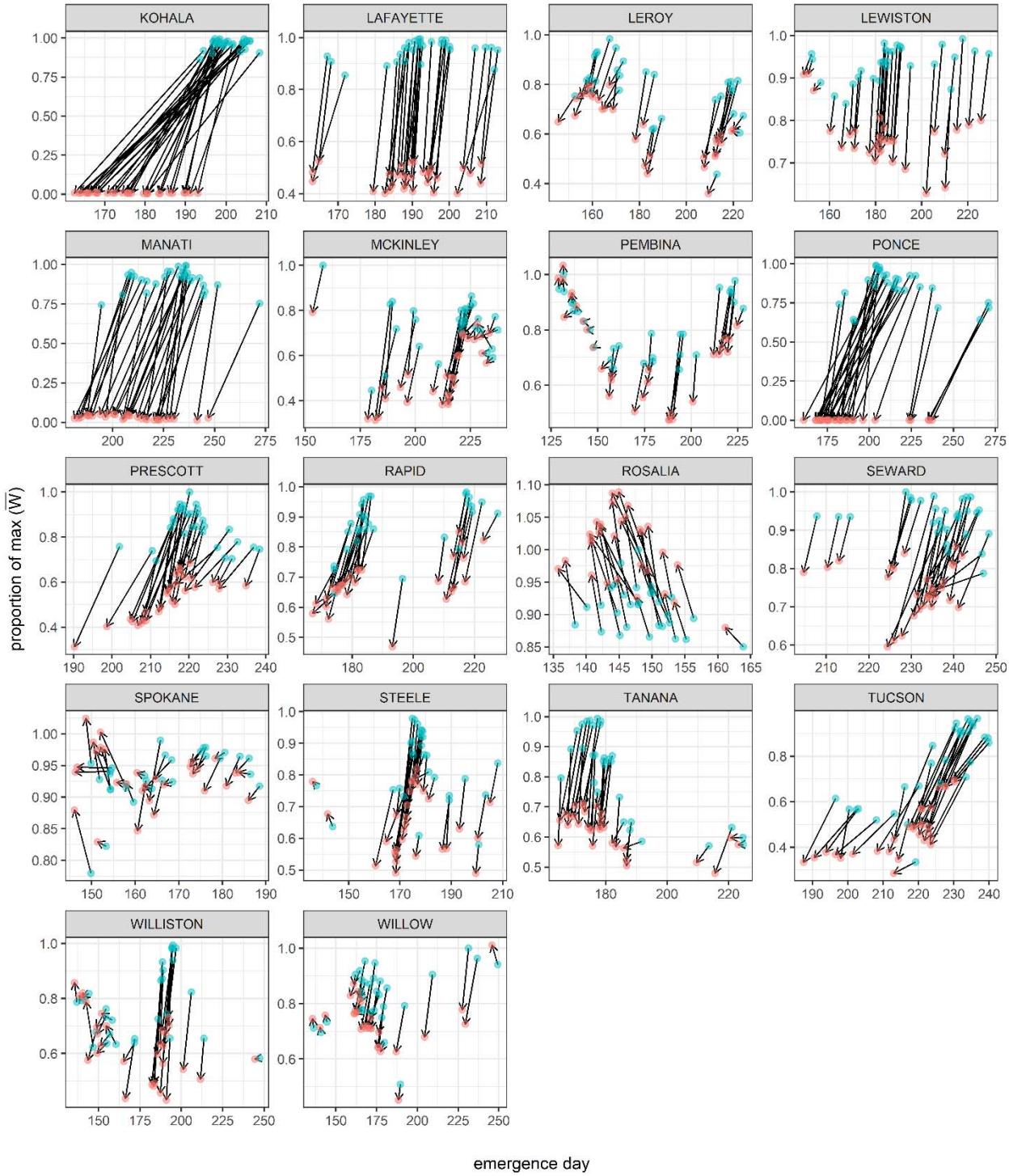
872

873


BUTTE
CALMAR
CARLSBAD
CARSON
COLOSO
CORDOVA
DAVIS
EASTPORT
FAIRMONT
FARMINGTON
FORT
GARDINER
GLENDIVE
GREAT
HAYS
HONOLULU
HORTON
ITHACA
KAILUA
KALISPELL
proportion of max (W)
emergence day

874

875


KOHALA
LAFAYETTE
LEROY
LEWISTON
MANATI
MCKINLEY
PEMBINA
PONCE
PRESCOTT
RAPID
ROSALIA
SEWARD
SPOKANE
STEELE
TANANA
TUCSON
WILLISTON
WILLOW
proportion of max (W)
emergence day

876

877

Supplementary Methods

Environmental Data

Most of the 78 locations were chosen haphazardly, with additional locations selected to ensure representative sites across the United States, several Atlantic and several Pacific island sites, and two sites in Canada. Locations chosen are listed in Supplementary Table 1 along with their ID for use in R data files. Imputation was conducted with the expectation-maximization with bootstrapping (EMB) algorithm implemented in Amelia II[1], with priors informed by long-term climatic daily means and standard deviations. Imputation was conducted with the script *"enviro_data_imputation_final.R"*. For each location, we calculated the minimum distance to an marine coastline using a high resolution coastal shapefile[2] with the *rdgal* package[3]. Coastal distance was calculated with the script *"distance-to-coast.R"*.

Sensitivity Analysis

We explored several alternative model structures, cues, and parameter values, including quadratic responses to temperature, precipitation, and day as well as daily – rather than cumulative – cues of temperature and precipitation. Using these variants produced qualitatively similar results with consistent spatial variation in cue use, but each additional cue has a multiplicative effect on potential trait space, and alternative models lacked a strong a priori basis.

Thorough exploration and examination of trait-space proved computationally prohibitive, but for each location we evaluated the geometric mean fitness of 5,000,000 genotypes generated from a latin hypercube implemented in lhs[4], constrained to the same "reasonable range" of traits we used for the population initialization in the simulations. During this evaluation, we examined how model results changed when we varied lag (the number of days between $E \geq 1$ and the organism emerging) between 1 and 5, and the temperature and moisture that maximized fitness between 5th and 95th percentile of observed cue values for each location. Results for these sensitivity analysis were qualitatively similar, but trait values and cue use changed when organisms are rewarded for emerging in low temperatures and moisture. We carried out test simulations in which organisms had different windows (1 and 20 days) after emergence to collect fitness. These simulations showed qualitatively different windows, although simulations with shorter windows were more influenced by extreme weather events, while simulations with longer windows provided smaller benefits for emerging on exactly the optimal day.

Climate Change Scenarios

We tested a "warming + shift" climate change scenario in addition to the two scenarios presented in the main manuscript. This third scenario included a 3-degree increase in daily temperatures, and a shift in the precipitation regime that matched the median phenological advancement in cumulative temperatures. To do this, we calculated the median cumulative temperature value for each location. For each year, we determined the day the cumulative temperature matched or exceeded this value in the historic climate regime and in the shifted temperature scenario. The difference of those two days represented the effective temperature advancement, and we shifted the precipitation regime forward to match this advancement, treating each year as cyclical to accommodate the beginning and end of each year. This produced qualitatively similar responses in the organisms, but this "warming + shift" scenario was complicated by unpredictable shifts in precipitation which varied by location (e.g., warm locations had smaller shifts than cold locations, since the relative change in cumulative temperature from a 3 degree warming was smaller).

Code Guide

The code is designed to work with the following file structure, which is necessary for the code to run. Files are italicized while folders are not. Exactly file names are contained in quotes. A zip file with the scripts and data files in the appropriate file structure can be found on Dryad.

- Project directory
  - data-years
    - *all data files to be used*
    - *"location summary with photoperiod.Rdata"*
  - data-years-lisovski
    - *all data files to be used*
  - fitcurve
    - *"multiskewgauss.R"*
  - parameters
    - Any parameter files to run. The following are the parameter files used to generate the results presented here:
    - *"cu-ccshift-mut005-parameters.R"*
    - *"cu-cctemp-mut005-parameters.R"*
    - *"cu-cuphotoperiod-mut005-parameters.R"*
  - results
  - scripts
    - *"master-script.R"*
    - *"rate-setup.R"*
    - *"simulation.R"*
    - *"windows-subs.R"*
  - yearinds

Code is initiated by running *master-script.R*. The "set_wrkdir()" function at the beginning of this file should be updated to point to the working directory, and "prefix" and "suffix" should be updated to match the name of the parameter file to run (by default, set "prefix" to everything before the ".R" of the parameter file name, and set "suffix" to be an empty string). The "sum.merge.only" and "test.only" parameters are to analyze existing runs and test the code using the first two locations, respectively. These should be set to FALSE for a complete run. Outside of the parameter files, these are the only lines of code that may require modification to implement our model.

After reading in parameter (below) and function files, *"master-script.R"* calls "*simulation.R"* script for every location. "*simulation.R"* reads in the imputed climate data files, and calculates derived climate measures (cumulative cues, fitness, etc). It then calls *"rate-setups.R"* to initialize mutation rates and starting trait ranges. *"simulation.R"* then uses the function "*year_var_analyze"* to calculate climate metrics, which are saved in "…-stats.Rdata" in the data-years folder. The script then saves the current scripts (into the results folder), parameters (into the results folder), and the random order of years to be used (into the yearinds folder). These are useful for tracking changes to the model as well as making comparisons between model runs. Climate change scenarios are then generated, creating a changed version of the locations climate data, which is saved in the results folder. *"simulation.R"* then uses parallel processing (using between 1 and 5 clusters depending on the computer's number of cores) to carry out the actual simulation, which is run through the *"run_sim()"* function. This and most other functions are defined in the "*windows-subs.R"* file. Results of each simulation run (along with intermediate objects and functions) are saved in the corresponding results folder in a file labeled *"dat.Rdata"*. The realized relative cue use is then calculated and saved in *"finalpop-yrtest-acteff-*

dat.Rdata" (see Methods for an explanation of this metric). As a reminder, this is calculated for each individual of the final population exposed to each year of the historic climate. The climate change scenario is then evaluated using the "*climate_master()"* function. This function takes the final population of the simulation and evaluates its emergence and performance on each year of the changed climate, saving the results in "*temp.Rdata*" in the climate folder within the corresponding results folder. Finally, fitness parameters used for this location are saved in "*…_fitparms.Rdata*" in the corresponding results folder.

After each simulations for each location have been run, "*master-script.R"* uses *"trait_eff_summarize_small()"*to aggregate the realized relative cue use information to the individual ("*acteff-agg-byindiv.Rdata"*) and the simulation level *("acteff-agg-bysim.Rdata"*) level in the corresponding results folder, and aggregates all simulation information to the simulation level in *"all-finalpop-yrtest-simlevel-….Rdata"* in the appropriate results folder. After all simulations and aggregations have been carried out, "*master-script.R"* calls *"dat_sum_merge"* to summarize climate metrics, and then merge these metrics with location data, simulation results, and realized relative cue use in historic and changed climate regimes, which is saved in into a single data frame with simulation-level summaries which is saved in *"climateVsPops-….Rdata"*. This file contains the majority of the information presented in this paper. Below is a description of each column's contents.

In these descriptions and in the parameter guide that follows, "simulation" refers to one of several identically parameterized instances of the model in a location; these are best imagined as alternate realities. The results we present had 30 simulations for each location for each set of parameters. "Runs" refers to different executions of the code (typically an execution of the code applies the parameters to every location, producing multiple simulations for each of 78 locations). Presumably these runs are carried out with the purpose of comparing consequences of differences in parameters; thus each run is likely to have a different parameter file.

The results file has a row for each simulation. Each simulation in a location had a different random sequence of years (1000 by default) drawn from the same original collection of years of recorded daily climate data with missing values imputed (see Environmental Data in Methods and Supplements). Unless stated otherwise, climate metrics were calculated based on the sequence of years actually experienced, producing slightly different values between simulations.

Results file column contents

- name: Location ID (see supplements table 1)
- sim: simulation number
- eff.day: simulation day effect, averaged across all individuals and all years using the *acomp* function in the *compositions* package (see realized relative cue use in Methods)
- eff.cutemp: as day, for cumulative temperature
- eff.cuprecip: as day, for cumulative precipitation
- emerge: emergence day of final population averaged across all years and all individuals
- emerge.cc: emergence day of final population under climate change scenario averaged across all years and all individuals
- geofit: raw units of "fitness" obtained by final generation (used to determine relative). Averaged across all individuals in all years
- geofit.cc: as geofit, but using climate change climate data
- lis.tpred: climate metric that represents the predictability of temperature[5]. This was calculated for the original sequence of years (for each location, the years were ordered chronologically and

treated as sequential). As such, the same value was used for all simulations with the same location.

- lis.ppred: as lis.tpred, but for precipitation
- lis.mpred: as lis.tpred, but for moisture
- lis.tseason: climate metric that represents the seasonality of temperature[5]
- lis.pseason: as lis.tseason, but for precipitation
- lis.mseason: as listseason, but for moisture
- pau.tvar: climate metric that represents variability in temperature[6]
- pau.mvar: as pau.tvar, but for moisture
- pau.pvar: as pau.tvar, but for precipitation
- pau.tseason: climate metric associated with each simulation. This metric represents seasonality of temperature[6]
- pau.mseason: as pau.tseason, but for moisture
- pau.pseason: as pau.tseason, but for precipitation
- b.day.mn: day trait of final population, averaged across all individuals
- b.cutemp.mn: as b.day.mn, but cumulative temperature
- b.cuprecip.mn: as b.day.mn, but cumulative precipitation
- med.cv: coefficient of variation in the "median day". For each year of the climate data, we determined the first day the cumulative temperature reached or exceeded half the maximum cumulative temperature of the mean year for that location. Coefficient of variation was calculated from these median days
- imput.temp.mn: fraction of the daily temperature values that originated form imputation
- imput.prec.mn: as imput.temp.mn, but for precipitation
- med.var: as med.cv, but variance instead of coefficieint of variation.
- within.mn: The "within-year unpredictability" was calculated for each year by comparing the daily temperature to the daily temperatures of the mean year. Each year was allowed to advance or retreat by discrete days, and the temperatures could increase or decrease by a constant; using sum of squared errors, we found the daily shift and temperature change constant to make the current year best match the average year. The "within-year unpredictability" was then the mean of the square of the lag 1 difference ("diff") of the residuals of this fit. This represents day-to-day inconsistencies that weren't present in the average year. This metric was calculated for each year of the climate, and then averaged across all years.
- within.nodiff.mn: as within.mn, but using the mean of the square of differences (skipping the "diff" step). This represents divergences between the current year and the average year.
- Var.dailyfit.mn: For each year, we calculated the variance in the relative fitness available on each day. This was averaged across all years.
- lr.var: The "left-right shift" for a year was the number of days advancement needed to best fit the current year's daily temperatures to the average year of that location. For each location, the lr.var was the variance of these measures across all years for that location.
- Updown.var: as lr.var, but using the temperature shift needed to best fit the current year's daily temperatures to the average year for that location.
- Temp.mn: mean of all yearly mean temps for the simulation (Note: this was calculated with the temperature values that had been shifted so that the minimum temperature in each location was 0)
- Temp.orig.mn: mean of all yearly mean temps for the simulation using the original, unshifted temperatures

- Temp.predtemp.mn: mean correlation of daily temperature on day n to sum of temperatures on days (n+1):(n+duration). As with all other X.predY.mn, this was a metric for how well X acted as a cue to predict the entire span of Y experienced by an individual that chose to emerge on day n.
- Temp.predfit.mn: as temp.predtemp.mn, but how well temperature predicted sum of fitness
- Temp.predmoist.mn: as temp.predtemp.mn, but how well temperature predicted sum of moisture
- Moist.mn: mean moisture across all years
- Moist.predmoist.mn: as temp.predtemp.mn, but how well moisture predicted sum of moisture
- Moist.predfit.mn: as temp.predtemp.mn, but how well moisture predicted sum of fitness
  Moist.predtemp.mn: as temp.predtemp.mn, but how well moisture predicted sum of temperature
- Moist.cv.mn: average across all years of the within-year coefficient of variation of daily moisture
- Precip.mn: average across all years of within-year mean precipitation
- Precip.var.btwn: between-year variance in the yearly mean precipitations
- Precip.var.mn: average across all years of within-year variance in precipitation
- Precip.cv.mn: as precip.var.mn, but using coefficient of variation
- Temp.varbtwn: between-year variance in the yearly mean temperatures
- Temp.cvbtwn: between-year coefficient of variation in the yearly mean temperatures
- Precip.cvbtwn: between-year coefficient of variation in the yearly mean precipitation
- Moist.varbtwn: between-year variance in the yearly mean moisture
- Moist.cvbtwn: between-year coefficient of variation in the yearly mean moisture
- Temp.lag1temp.mn: mean lag 1 autocorrelation of daily temperature. As with all X.lag1Y.mn metrics, this was a simple metric for how well X acted as cue to predict Y
- Temp.lag1moist.mn: mean correlation between temperature of day n and moisture of day n+1
- Precip.lag1precip.mn: mean lag 1 autocorrelation of precipitation
- Moist.lag1temp.mn: mean correlation between moisture on day n and temperature on day n+1
- Moist.lag1moist.mn: mean lag1 autocorrelation of moisture
- Moist.mn.var: same as moist.varbtwn
- Temp.cv.mn: mean across years of the within-year coefficient of variation of daily temperatures
- Temp.varwin.mn: as temp.cv.mean, but variance instead of coefficient of variation
- Moist.varwin.mn: as temp.varwin.mn, but moisture instead of temperature
- Lat: latitude of location
- Lon: longitude of location
- Elev: elevation of location
- Coast.dist: distance from location to nearest coast
- Count.good: number of years of climate data available.

Parameter file

The specifics of any simulation run are controlled by the parameters defined in the parameter file. Below are descriptions of each parameter

- runs.types: the name of each location file to use, in vector of characters
- traits: used for sensitivity analysis. Changing from “day”, “cutemp”, and “cuprecip” may cause the current code to break
- num.sims (default: 30): the number of simulations to run for each location. Each simulation has identical parameterizations, but the randomly generated starting genotypes, randomly generated sequence of years, and randomly generated mutations of each generation will differ.

- N (default: 500): number of individuals per simulation
- num.years (default 1000): number of years to simulate.
- duration (default: 10): number of days organism gathers fitness
- lag (default: 1): number of days after organism decides to emerge before it begins gathering fitness
- mut.dist (default: .005): fraction of trait range to use as standard deviation of mutation rate
- fattail (default: FALSE): indicator to determine whether to use a normal distribution or Cauchy distribution to generate mutation values. Early testing showed that the two distributions produced indistinguishable dynamics in the model, and subsequent simulations used the normal distribution. Setting this to TRUE may cause the current code to break.
- base.temp (default: 0): threshold below which temperature isn't added to cumulative temperature
- decay (default: .8): decay parameter alpha for use when calculating moisture (see Methods)
- fit.shape (default: "skewgauss"): defines which fitness shape script to use. The "*skewgauss.R*" is the only fitness shape script designed to work with the current code.
- Other.name (default: "moist"): determines what daily environmental measure combines with temperature to define fitness.
- shape.temp (default: 15): initial guess at shape parameter for the temperature fitness curve. The final shape parameter is determined by minimizing sum of squared errors.
- shape.other (default: 15): as shape.temp, but for the other fitness-determining environmental measure (by default, moisture)
- best.temp.quant (default: .9): Used to fit fitness curves. At this quantile of temperatures for each location, the skew normal has its maximum value (see Methods)
- best.other.quant (default: 0.9): as best.temp.quant, but for the other measure (by default, moisture)
- min.quant (default: 0.1): Used to fit fitness curves. At this quantile of temperature and moisture, skew normal function has a value that is (ratio.min) of the peak. (see Methods)
- ratio.min (default: 0.1): determines ratio of values at the min.quant value and best.*.quant value for each skew normal curve.
- plot.extra (default: FALSE): If true, plot emergence of a subset of the generations of each simulation
- plot.pheno (default: FALSE): If true, plot phenotype of a subset of the generations of each simulation
- burnin (default: 100): Used for emergence and phenotype plots, defines how many of the initial generations to skip in the plots
- year.label (default: "A"): When the simulation script is generating a random sequence of years for all simulations in a given location, it checks to see if there exists a file storing random sequences for this location with the same num.sims, num.years, year.set (below) and year.label; if it finds such a file, it uses those randomly generated year sequences. That is, no two simulations within a run will have the same sequence of years, but using the same year.label, num.sims, num.years, and year.set values when running two different parameter files will re-use the same num.sim (default: 30) randomly generated sequences of years for each of the separate parameter file runs. This allows for fair comparisons between different model structures or parameterizations (since ABERDEEN.MS run 1 of each parameterization will have the same sequence of years). Changing the year.label between parameter files causes each to have a separate sequence of years.
- year.set (default: "all.years"): as year.label – shows up in a different part of the naming scheme but plays the same role.

- temp.increase (default: 0 for “shift” climate change, 3 for “warming” climate change): how much should each daily temperature be increased in the climate change scenario?
- precip.shift (default: 5 for “shift” climate change, 0 for “warming climate change”): how many days should precipitation be advanced in the climate change scenario?
- Lockstep (default: TRUE): should temperature advance with precipitation in the climate change scenario?
- tempsd.increase (default: 1.2): for use in deprecated climate change scenario
- phensd.increase (default: 1.2): for use in deprecated climate change scenario
- intra.increase (default: 1.2): for use in deprecated climate change scenario
- scen.temp (default: TRUE): with change in climate change functions, this should always be TRUE
- scen.tempsd (default: FALSE): with change in climate change functions, this should always be FALSE
- scen.phensd (default: FALSE): with change in climate change functions, this should always be FALSE
- scen.intra (default: FALSE): with change in climate change functions, this should always be FALSE
- save.small (default: TRUE): If FALSE, saves information on every individual of every generation in each simulation. If TRUE, only saves information on every individual of the first and final generations.
- photo.flag (default: FALSE): if TRUE, use cumulative photoperiod instead of day of year for the “day” cue.
- do.lisovski (default: FALSE): if TRUE, calculates metrics based on Lisovski et al.[5] , and saves them in separate data files in data-years-lisovski. As this is a lengthy calculation, it’s best to re-use existing calculations of this (and the data files of these calculations can be found in **X**).
- stats.lisovski: if TRUE, uses the metrics associated with do.lisovski. This should be TRUE unless you (a) are missing the lisovski metric data files, (b) don’t want to wait the additional hours for the scripts to calculate them, and (c) don’t want to use those metrics.

References

1. Honaker, J., King, G. & Blackwell, M. Amelia II: A Program for Missing Data. *J. Stat. Softw.* **45,** 1–47 (2011).
2. National Geophysical Data Center. Shoreline/Coastline Databases | NCEI. Available at: https://www.ngdc.noaa.gov/mgg/shorelines/gshhs.html. (Accessed: 20th September 2018)
3. Bivand, R., Keitt, T. & Rowlingson, B. *rgdal: Bindings for the ‘Geospatial’ Data Abstraction Library*. (2018).
4. Carnell, R. *lhs: Latin Hypercube Samples.* (2018).
5. Lisovski, S., Ramenofsky, M. & Wingfield, J. C. Defining the degree of seasonality and its significance for future research. *Integr. Comp. Biol.* icx040–icx040 (2017). doi:10.1093/icb/icx040

6. Pau, S. *et al*. Predicting phenology by integrating ecology, evolution and climate science. *Glob. Change Biol.* **17,** 3633–3643 (2011).