Evidence for continent-wide convergent evolution and stasis throughout 150 years of a biological invasion

Yihan Wu[1] and Robert I. Colautti
Department of Biology, Queen's University, Kingston, Ontario, Canada
[1]Current Address: Botany Department, University of British Columbia, Vancouver, BC, Canada

*Corresponding authors Robert I. Colautti and Yihan Wu

Email: robert.colautti@queensu.ca; yihan.wu@botany.ubc.ca

Author Contributions: Both authors contributed to conceptualization, analysis and writing. In addition, YW wrote the final code and RIC was responsible for funding, supervision and project administration.

Competing Interest Statement: The authors report no competing interests.

Classification: Biological Sciences/Ecology

Keywords: *Lythrum salicaria*; Virtual Common Garden (VCG); herbarium records; punctuated equilibrium; invasive species; contemporary evolution

Abstract

The extent to which evolution can rescue a species from extinction, or facilitate range expansion, depends critically on the rate, duration, and geographical extent of the evolutionary response to natural selection. Adaptive evolution can occur quickly, but the duration and geographical extent of contemporary evolution in natural systems remains poorly studied. This is particularly true for species with large geographical ranges and for timescales that lie between 'long-term' field experiments and the fossil record. Here, we introduce the Virtual Common Garden (VCG) to investigate phenotypic evolution in natural history collections while controlling for phenotypic plasticity in response to local growing conditions. Reconstructing 150 years of evolution in *Lythrum salicaria* (purple loosestrife) as it invaded North America, we analyze phenology measurements of 3,429 herbarium records, reconstruct growing conditions from more than 12 million local temperature records, and validate predictions across three common gardens spanning 10 degrees of latitude. We find that phenology clines have evolved along parallel climatic gradients, repeatedly throughout the range, during the first century of evolution. Thereafter, the rate of microevolution stalls, recapitulating macroevolutionary stasis observed in the fossil record. Our study demonstrates that preserved specimens are a critical resource for investigating limits to evolution in natural populations. Our results show how natural selection and trade-offs measured in field studies predict adaptive divergence observable in herbarium specimens over 15 decades at a continental scale.

Introduction

Global biodiversity in the Anthropocene is threatened by the jointly homogenizing effects of population extirpation and biological invasion (1). These outcomes lie along a spectrum of population growth trajectories that depend fundamentally on environmental (mal)adaptation (2). Adaptive evolution in novel and changing environments can rescue populations from extinction (3) and facilitate the spread of invasive species (4). However, genetic constraints can limit the evolutionary response to selection such that the balance of adaptation and constraint within a species determine its ecological niche (5). Species distribution models and management strategies under current and future global change could benefit from a better understanding of the balance between natural selection and genetic constraint, and how this balance affects the rate, duration, and geographical extent of adaptive evolution across a species range.

Selection gradients and rates of phenotypic evolution have been studied in native populations for a variety of taxa (6–9). In contrast, little is known about selection operating in invasive populations, partly due to an apparent researcher bias against measuring selection in non-equilibrial populations (8). More generally, the duration and geographical extent of adaptive evolution are difficult to assess in field experiments, especially when one is interested in variation throughout a species range over decadal to centenary timescales. Common garden studies involving genotypes from contemporary populations can offer insight into local adaptation, but provide only a snapshot of evolutionary change (10–14). 'Long-term' ecological studies can reveal temporal dynamics that are not evident in short-term experiments, but are rare due to considerable logistic challenges that also limit the extent of spatial replication (15, 16). Therefore, it remains unclear how the rate and duration of adaptive evolution vary throughout species ranges, and whether evolution can be sustained over the decade to century timescales most relevant to conserve and manage biodiversity for the 21st century.

Invasive species provide opportunities to study evolution in action, but inferences drawn from the genetics of contemporary populations are complicated by demography (i.e. invasion history) and spatio-temporal variation in natural selection (8). Clines in growth and phenology have been identified in many invasive plants, but the rate, duration, and adaptive significance of cline evolution remains uncertain (8, 12). Whether these latitudinal clines are adaptive remains largely untested, but a decade of research has shown that flowering time clines are genetically-based and locally adaptive in invasive populations of *Lythrum salicaria* (purple loosestrife) from eastern North America (17, 18). Local adaptation in these populations likely evolved as a consequence of (i) a trade-off between growth and phenology that is consistent across growing environments, and (ii) directional selection on flowering time with steeper gradients in shorter growing seasons (18–21). This research demonstrates adaptive cline evolution during invasion of *L. salicaria* but covers a relatively small portion of the North American distribution. It is therefore uncertain whether parallel clines evolved elsewhere in North America, and how quickly these adaptive clines evolved during invasion.

Adaptive clines could evolve along two distinct trajectories analogous to long-term evolutionary dynamics observed in the fossil record (Fig. 1). First, a Continuous Evolution Model (CEM) predicts a gradual and relatively constant rate of phenotypic change through time (Fig. 1a,b). Alternatively, a punctuated equilibrium model (PEM) predicts an early, rapid burst of evolution

followed by a static equilibrium (22). These scenarios represent opposite ends of a spectrum defined by (i) the strength of natural selection and (ii) the amount of genetic variation for the trait under selection (23). In the CEM, evolution is relatively slow and approximately continuous because the change in fitness across environments is approximately linear and genetic variation is not limiting. In the PEM, a punctuated response to evolution occurs when there is strong directional selection initially, but populations quickly reach their new optima and/or standing genetic variation is depleted during the response to selection. Distinguishing which models apply to natural populations is important for understanding the rate of contemporary evolution in response to novel and changing environments. Yet, testing these alternatives across a species range requires quantifying rates of evolutionary change at multiple locations.

Temporal sampling of genetic changes in natural populations provides some of the strongest evidence for adaptive evolution in nature (16, 24), but these experiments are difficult to replicate over large spatial and temporal scales (25–27). In contrast, millions of preserved specimens maintained in global natural history collections offer detailed records of phenotypes sampled over decades to centuries past, holding potential for reconstructing evolutionary trajectories (28–30). Indeed, herbarium records can be used to reliably reconstruct phenology variation over space and time (31), but evolutionary shifts in phenotypic traits observed under natural field conditions are difficult to distinguish from developmental plasticity associated with different environments (32). Thus, it would be valuable to develop a mechanistic model, informed by field experiments, to disentangle phenotypic plasticity from genetic effects on phenotypes observed in natural history collections. Such a model could enable tests of adaptive evolution and genetic constraint across broad temporal and spatial scales unachievable by other methods.

The virtual common garden (VCG) is a computational approach that incorporates biophysical models informed by field experiments to reconstruct phenotypic plasticity and genetic differentiation over fifteen decades at a continental scale. As such, the VCG is a type of observational experiment for eco-evolutionary inference. Using this approach, we reconstruct 150 years of phenotypic evolution in the self-incompatible wetland perennial *Lythrum salicaria* (purple loosestrife) as it spread across North America. Our specific goals are: (i) to parameterize and then validate the VCG as a method to distinguish genetic variation from developmental plasticity in *L. salicaria*; (ii) to test for convergent evolution via local adaptation across North America; (iii) to test the CEM and PEM models shown in Figure 1 by measuring changes in the rate of cline evolution through time.

Results

Herbarium Specimens

We calculated the spatially-weighted interpolation of Growing Degree Days up to the date of specimen collection (GDD_C) and the total growing Season Length (SL) for each of the 3,429 digitized herbarium specimens. These were based on >12 million temperature records, used to reconstruct 62,208 annual temperature accumulation curves from 6,303 weather stations (Fig. 2).

Measurements from all digitized herbarium specimens that met inclusion criteria (Fig. 2) are available online in the Dryad Database (https://doi.org/10.5061/dryad.qfttdz0f5) with images available from sources cited therein. The earliest specimen was collected on July 29th, 1866

near Boston, Massachusetts, with the number of samples increasing thereafter to a peak in the 1970s and declining in recent decades (Fig. S2). The most recent specimen was collected on August 31, 2016 from Lake Lowell in Idaho. Specimens cover a large part of North America spanning from Atlantic to Pacific coasts, and almost 20 degrees of latitude, from as far south as Carthage, Mississippi, USA (32.74°N) to Prince George in BC, Canada in the north (54°N). A majority of specimens were collected in the East Coast and Midwest regions (Fig. 2), each with over 1,000 specimens, while the West region contains approximately 400 specimens (Table S1). The most densely sampled regions are the Great Lakes region (particularly in Illinois and Wisconsin), along the St. Lawrence River, and along the northeastern coastal region of North America (Fig. 2).

The average Julian day of collection was 217 d (*sd* = 25.18), or August 5 in a non-leap year, and ranged from 121 d (May 1, 1938) to 365 d (December 31, 1906). This represents a broad range of days measured from the start of the growing season at each location, with a mean of 98 d (min = 8, max 263) (Fig. S4). Growing season length (SL) estimated from weather stations varied across regions, with the average growing season longer in the West Coast region and decreasing eastward; SL was also strongly correlated to latitude (r = -0.6). Specimens captured a broad distribution of phenological stages from a few buds to fully mature fruits. There was no clear evidence for bias with respect to phenology sampling, as all but the earliest phenologies were evenly distributed (Fig. S3). Nor did we find strong clustering of phenologies with respect to geography, even though some structure is expected due to the predicted relationship between season length and latitude (Fig. S5). However, we did observe variation in sampling intensity across the range (Fig. 1), which we address using spatially-explicit resampling methods as described later.

VCG Validation

Details of the VCG analysis are provided in the Materials and Methods, Supplementary Methods and online repository (DOI) (https://doi.org/10.5061/dryad.qfttdz0f5). The VCG analysis involves several key steps, which are outlined in the Materials and Methods section.

To validate the VCG we compared (i) latitudinal clines in development time (ψ) of herbarium specimens with (ii) days to first flower in three real-world common garden experiments in eastern North America (data from (18)), each standardized to a mean of zero and unit variance. This comparison involves heterogenous phenology measurements (i.e. ψ and days to first flower) and growing conditions (i.e. a virtual common garden and three 'real world' garden sites spanning 10° of latitude), yet the estimated latitudinal clines were remarkably consistent (Fig. 3). The correlation coefficients (i.e. clines) observed in the VCG (r = -0.70, bootstrapped 95% CI: -0.17 to -0.92) was not significantly different from those measured in real-world field transplant sites at northern (r = -0.84, CI: -0.70 to -0.95), mid-latitude (r = -0.77, CI: -0.60 to -0.90) and southern gardens (r = -0.87, CI: -0.75 to -0.95).

VCG Cline Evolution

Clines in phenology were observed in all three geographic regions for specimens collected after 1980, and generally weaken back through time (Fig. 4, Table S3). East Coast populations were already well-established long before the first specimen collection in 1866 (33), and show a relatively stable cline until the present day. In contrast, contemporary populations in the Midwest have a similar cline that weakens back toward the time of establishment (~1900), while

West Coast populations did not establish until the 1920s and have not reached the same magnitude of cline as the two older regions (Fig. 4).

As a formal statistical test for a change in evolutionary rates across North America, we first modelled phenological development as a sigmoidal function of accumulated growing degree days (GDD_c). Analogous to the traditional common garden experiment, this model attempts to minimize phenotypic variance associated with developmental plasticity. To test for local adaptation, we add a parameter for season length (*SL*) as the predicted genetic correlation between season length and phenology. Parameter estimates for GDD_c and *SL* were significantly different from zero, and there was a significant $GDD_c \times SL$ interaction (Table S4). Together, these parameters represent a shift in phenology curves (Fig. 5a) with southern genotypes (red curves) predicted to take longer (x-axis) to reach a particular phenological stage (y-axis). If days to first flower is mapped to the x- and y-axes in Fig. 5a, the model recovers the common garden clines shown in Figure 3 in which genotypes from southern locales with longer growing seasons flower later than genotypes from shorter growing seasons.

In addition to testing for phenology clines, we used the same model to estimate evolutionary rates. Using a spatially explicit bootstrap resampling model to estimate *SL* and *GDD*c parameters over time, we found that the inferred rate of evolution was significantly different from zero (non-zero slope), and decelerating (significant quadratic term), rather than remaining constant (linear slope) during invasion (Fig 5c). Specifically, the coefficient estimates for linear and quadratic models are reported in Table S5 for the effect of temperature accumulation on phenology ($\beta_{b,t}$), and Table S6 shows the coefficients for the effect of season length on phenology ($\beta_{d,t}$). Based on Likelihood Ratio Tests (LRT), the quadratic terms were significant for both $\beta_{b,t}$ (χ^2 = 5448, p < 0.001) and $\beta_{d,t}$ (χ^2 =3045, p < 0.001), with a ~15% increase in R^2 values ($\beta_{b,t}$: 0.763 to 0.876; β_d: 0.713 to 0.810). This is shown in Figure 5 wherein the coefficients for $\beta_{b,t}$ (Fig. 5b) and β_d (Fig. 5c) stabilize around 100 years post-establishment.

Discussion

Geographical clines in morphological, physiological, and life-history traits are common to many taxa, but it has been difficult to measure the rate, duration, and geographical extent of cline evolution occurring in natural populations. In this study, we introduced the Virtual Common Garden (VCG) analysis as a multi-step computational approach to test evolutionary hypotheses using natural history collections. Below, we explore the relevance of our findings to the spread of invasive species and more generally to predict adaptive evolution in novel and changing environments.

Convergent Evolution During Invasion

Feedbacks between colonization and trait evolution can facilitate rapid range expansion (4, 34, 35), but clines *per se* do not provide clear evidence for adaptive evolution. In particular, simulations show how non-adaptive and maladaptive clines may evolve frequently when repeated founder events occur during spread, producing statistically significant but random slopes throughout the range (8, 36). In the case of *L. salicaria*, reciprocal transplant experiments spanning > 1,000 km and five growing seasons confirm that populations are locally adapted in eastern North America, with adaptive phenologies boosting reproductive fitness by one to two orders of magnitude (18). Heretofore, it had been unknown whether natural selection was likely

a causal agent of cline formation or merely acted in a supporting role, maintaining clines that evolved primarily through stochastic processes. Our VCG results do not support the prediction of random clines produced by founder effects and drift. Rather, similar clines evolved repeatedly throughout the range and strengthened over time (Fig. 4), implicating adaptive evolution as an important driver of phenology evolution across North America. These replicate, adaptive clines offer a case-study in intraspecific convergent evolution of a quantitative trait.

Classic examples of convergent evolution typically focus on distantly related species with similar but derived adaptations. In contrast, intraspecific convergent evolution of a phenotypic trait is not well-defined (37–39). In our analysis, phenologies of geographically isolated populations converge when they experience similar season lengths, while adjacent populations diverge when season lengths differ (Figs. 4 and 5). Regardless of the relative contribution of ancestral and derived alleles, adaptive evolution across a heterogenous adaptive landscape should leave a signature of phenotypic convergence, which we observe in the herbarium record as a reduction in spatial entropy of ψ over time. Intraspecific convergent evolution via local adaptation has implications for predicting the fate of introduced and native species in the Anthropocene.

Adaptation, Constraint and Stasis

Rapid evolution can facilitate persistence and spread in novel and changing environments, but trade-offs may constrain the rate and duration of adaptive evolution. We therefore hypothesized that adaptive clines could evolve along two distinct trajectories analogous to long-term evolutionary dynamics observed in the fossil record (Fig. 1). Our analysis supports the Punctuated Equilibrium Model (PEM) over the Continuous Evolution Model (CEM), with a rapid but relatively constant rate of cline evolution for ~100 years followed by a contemporary period of relative stasis (Fig. 5). Although several different scenarios of natural selection and genetic variation could produce a decline in evolutionary rates consistent with the PEM (Figs. 1 and 5), previous studies of *L. salicaria* suggest trade-offs as a primary factor.

Multiple common garden studies of *L. salicaria* from eastern North America reveal ample genetic variation within and among populations, measured for both molecular markers and quantitative traits (17, 20, 40). Populations also experience strong directional selection for flowering time, growth, and related quantitative traits (18, 19). Yet, despite significant standing genetic variation and strong directional selection, local adaptation is maintained by antagonistic selection on correlated traits involved in growth and phenology (18, 21). A punctuated evolution of clines followed by the stabilization of evolutionary rates consistent with PEM add evidence that this adaptation-constraint phenomenon may be relevant to adaptation throughout a species range. Specifically, adaptive evolution should be fastest in the early stages of colonization when newly established populations are far from their phenotypic optima and thus experience stronger directional selection (Fig. 1). Evolutionary rates should decline thereafter as antagonistic selection on plant size maintains the local optima, resulting in a static equilibrium that is not observed for the first several decades of invasion. It is an open question as to whether this phenomenon is common among biological invasions and more generally whether rates of evolution measured in contemporary populations will stall over longer timescales.

An inherent assumption in the application of a VCG analysis is that plasticity of a trait of interest (e.g., phenological development time) can be predicted by local growing conditions (e.g., GDD_c), with limited genotype-by-environment interactions (G x E). Another limitation is that the model

cannot parse unmeasured environmental characteristics (U_N) (e.g., photoperiod), which could confound our interpretation if the U_N affect growth and correlate with *SL*, but do not similarly correlate with GDD_c. Multiple common garden experiments with *L. salicaria* confirm that clines in growth and phenology are similar across rearing environments with limited influence of G × E or photoperiod effects (17–19, 41). Moreover, confounding effects of G x E or U_N would have been evident as non-parallel clines in the common garden data shown in Figure 3. Instead, we observe a constancy of clines across rearing environments, both real and virtual.

A pattern of rapid evolution and stasis may be common during invasion, given that the climatic niche is often conserved between the native and introduced range (42). Similar to *L. salicaria*, many other invasive species also (i) exhibit genetically-based clines, (ii) experience directional selection, and (iii) segregate genetic variation for quantitative traits at levels comparable to genotypes from their native range (8, 12). It is unclear whether native species would show similar dynamics during natural colonization (e.g. post-glaciation), as the genetics of native populations can be quite different (2). Nevertheless, our study demonstrates how genetic constraints can help predict evolutionary outcomes that may take decades to manifest. A key goal for future VCG studies is to identify genetic clines and differentiate those that evolved through stochastic processes (e.g. random clines) vs. natural selection (e.g. convergence). Convergent cline evolution and stasis could provide a foundation for investigating constraints on evolution in novel environments that might not be evident for the first few decades of evolution.

Natural History Collections and Evolution

Studies of evolutionary change in natural populations often rely on demographic inferences from molecular markers, but inferences from allele frequency data are known to be problematic for genetic loci of quantitative traits like phenology, and for populations with complex demographic histories that have not reached an equilibrium between migration/mutation and drift/selection (8, 43, 44). For example, a shift from strong directional to weak stabilizing selection on a set of multi-locus traits following the PEM scenario in Figure 1 could leave a genomic signature similar to a demographic bottleneck followed by population expansion. These limitations help to explain why Q_{ST}-F_{ST} analyses of *L. salicaria* populations do not reveal selection on growth and phenology (40), contrary to clear evidence in field studies that these traits are under strong selection (18, 19, 21). In contrast, our VCG analysis identified convergent clines that are consistent with locally adaptive clines confirmed in field experiments (Fig. 3). Additionally, the VCG revealed evidence for repeated evolution of clines across decades and thousands of kilometers (Fig. 4), with variable rates of change (Fig. 5). Although inferences from neutral genetic markers in contemporary populations may be limited, the application of molecular markers in combination with herbarium records could help to overcome some of these inherent limitations.

Informed by prior empirical studies and trade-off models, convergent cline evolution and stasis inferred from herbarium records of *L. salicaria* demonstrate how evolution of ecologically important traits in natural systems can be predictable on scales most relevant to modern human civilization: continental to global spatial scales and decadal to centenary timescales. A VCG analysis of herbarium specimens is just one example of how natural history collections can be mobilized to complement limitations of scale and simplism necessary for tractable experimentation. Preserved specimens offer a rare opportunity to rewind the clock and

reconstruct phenotypic change, but they are vulnerable to loss without significant financial and academic investment. Global efforts to manage biodiversity in the Anthropocene could benefit from a more predictive science of adaptive evolution and constraint in natural systems, made possible only with the support of natural history collections at a global scale.

Materials and Methods

Virtual Common Garden

Common garden studies are commonly used to identify and quantify genetic differences in quantitative traits. Genetic variation is inferred from phenotypic differences among populations, genetic families, or genetic clones when all individuals are reared under similar growing conditions. The Virtual Common Garden (VCG) applies the same philosophy but uses a computational approach that includes the following steps:

1. Convert inflorescence measurements of each specimen to a size-independent and scale-free index of phenological stage (φ).
2. Generate histograms of φ to look for sampling bias.
3. Map the geographic location and phenological stage of each specimen to look for spatial bias.
4. For each specimen, download weather data for up to 20 nearby weather stations.
5. For each weather station, calculate a vector of 365 (or 366) average daily temperatures for the same year in which a nearby specimen was collected.
6. Interpolate a vector of average daily temperature for each specimen by inverse distance weighting.
7. Estimate the start and end of the growing season for each specimen to define the local season length (SL).
8. Calculate the local growing-degree days from the start of the growing season to the day of collection (GDD_c) for each specimen.
9. Fit a statistical model of observed phenological stage (φ) and collection date to estimate a phenological development time (ψ) for each specimen.
10. Use spatial Kriging to interpolate population time since establishment for each specimen (Age).

Using these data, the VCG does the following:

11. Validate ψ by comparison with flowering times observed in common garden experiments.
12. Calculate correlation between ψ and season length, and test whether these change over space and time.
13. Use another statistical model that will (i) predict the joint effects of GDD_c and local adaptation to season length, and (ii) test whether these relationships (i.e. model coefficients) change over time.
14. Resample, with replacement, individual specimens at 0.1 ° latitude/longitude resolution to generate bootstrap means and confidence intervals for the model parameters.

Phenology of Herbarium Specimens

Between August 2016 and December 2017, we analyzed 3,429 digitized herbarium specimens obtained from five sources: (i) the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/), (ii) the Regional Networks of North America Herbaria accessed through the Arizona-Mexico Chapter (http://swbiodiversity.org), (iii) the New York Botanical Garden (NYBG, http://www.nybg.org/), and (iv) the Database of Vascular Plants of Canada (VASCAN) (*20*)(45) and v) correspondence with 19 university herbaria to obtain more images (see acknowledgements). We included only specimens with both a full collection date (year, month, day) and location information that could be georeferenced (see Supplementary Methods).

Inflorescence development in *L. salicaria* occurs acropetally from the base of the stem to the tip of the apical meristem, resulting in three distinct regions of the inflorescence: fruits (basal), flowers, and buds (distal). To capture variation in floral phenology, we measured the length of each segment on each specimen using the segmented tool in ImageJ (46). We did this for the primary meristem, unless it was clearly damaged in which case the longest inflorescence was measured. Unpollinated flowers occur early in the phenology when compatible mates or pollinators are unavailable. As a result, they senesce and fall from the stem, leaving distinctive callouses at the point of attachment that represent the same phenological stage as fully-formed fruits. Using these measurements, we calculated a scale-free phenological stage (φ) for each herbarium specimen i:

$$\varphi_i = \frac{0 * buds_i + 0.5 * flowers_i + 1 * fruits_i}{total_i}$$

Values of φ range from 0 (early) to 1 (late) as a measure of phenological stage at the time of collection. Thus, on any given collection date, specimens with a phenology stage closer to 0 represent phenotypes early in their phenology whereas those with a stage closer to 1 represent phenotypes sampled later in phenological development.

Herbarium specimens are not often random samples of natural populations. Instead, sampling locations are chosen deliberately or opportunistically, introducing potential sampling biases. For example, first flowering dates observed in herbarium specimens collected from New England (USA) were biased toward a delay of ~3d relative to field observations; however dates were highly correlated with field observations despite this bias (31). Spatial and temporal clustering of samples is also likely when multiple specimens are often collected in sampling excursions conducted by a single individual or group. We therefore developed an analysis that would be robust to phenological stage, sampling date, absolute flowering date, and spatial clustering, as described below.

Season Length and Growth Conditions

Season length (*SL*) is a key predictor of genetic differentiation for flowering phenology in *L. salicaria*, whereas growing-degree days at the time of collection (GDD_C) determines the rate of growth and development and thus plasticity in phenology (18, 47–49). Both *SL* and GDD_C were interpolated from computation of daily weather data available from the Global Historical Climatology Network-Daily database (50). The algorithms and code for the computational interpolation are available on the Dryad Database (https://doi.org/10.5061/dryad.qfttdz0f5). Briefly, we collected temperature data for each herbarium specimen from up to 20 of the closest weather stations located within 0.5 ° latitude and longitude (Fig. 2). For each weather

station, we identified the first interval of at least 10 consecutive days above a threshold growing temperature (8 °C) from Jan 1 in the year of collection. The first day of this interval was set as the start of the growing season, and the end of the growing season was determined as the first day thereafter in which the temperature fell below 8° C. Temperature accumulation above 8° C is a key factor affecting growth and development of *L. salicaria*, and commonly used in growth models for the species (17, 47, 49, 51). A length of 10 days at the start of the season was chosen to avoid shorter intervals of abnormally warm winter or spring temperatures where growth would be unlikely. We then used the season start date to calculate the cumulative growing degree days until the day of collection (GDD_C). In summary, GDD_C characterizes the local growing environment in order to correct for variation in growing conditions and sampling date, whereas variation in season length (*SL*) is predicted to drive differentiation via local adaptation to climate.

Phenological Development Time

Field surveys and common garden experiments with *L. salicaria* have demonstrated both plastic and genetic variation for flowering phenology – warmer temperatures accelerate phenological development, whereas natural selection has caused the evolution of earlier flowering in short growing seasons (18, 47). By comparison, population-by-environment interactions are relatively weak, even across field sites spanning a latitudinal gradient of 1,000 km (18). Leveraging these observations, the VCG uses temperature accumulation data to model phenology as the additive effects of genotype and environment (plasticity). Since we cannot retroactively grow herbarium specimens in a common garden, we instead correct the observed phenological stage of each specimen (φ_i) for variation in the local growing environment (E_i) to estimate a relative measure of phenological development time due to genetic factors (ψ_i):

[Eq. 1] $$\psi_i = E_i - \varphi_i$$

As described later, E_i is the predicted contribution of environment to phenotype in a nonlinear least-squares model of phenology as a function of GDD_C. The equation for phenological development time (ψ) is therefore the residual phenological stage after accounting for GDD_C, multiplied by negative one. We use the negative multiple so that values for phenology in herbarium specimens lay in the same direction of more common phenological measurements (e.g. days to first flower, time to maturity).

In summary, we use two indices to describe the phenology of each herbarium specimen: The phenological stage (φ) is calculated directly from inflorescence measurements. The phenological development time (ψ) is an estimate of relative phenology after controlling for local growing environment. Phenological stage (φ) ranges from 0 (early stage) to 1 (late stage) and is determined by genetic factors, local growing conditions, and collection date. In contrast, the phenological development time (ψ) ranges from -1 (fast phenology) to 1 (slow phenology) and predicts genetic effects on phenology. As such, ψ is a standardized phenology metric that predicts which phenotypes would be observed when grown under similar rearing conditions – a virtual common garden.

VCG Validation

To validate the VCG, we compare the predicted phenological development time (ψ) inferred from herbarium specimens with observed flowering time clines reported in three separate common garden field experiments. These experiments are part of a previously published

reciprocal transplant study involving six populations sampled along a gradient of 10˚ latitude in eastern North America and grown at three sites: Timmins, Ontario (TIM), the Koffler Scientific Reserve (KSR) near Newmarket, Ontario, Canada, and the Blandy Experimental Farm (BEF) near Boyce, Virginia (18).

To facilitate comparison of virtual and real common garden metrics, we standardized population means within each dataset to z-scores (Mean = 0, SD = 1) and included only herbarium specimens collected after 1960 from the same geographic range in northeastern North America. Note that we standardize population means, rather than individual specimens, because population means are the unit of comparison between herbarium samples and common garden plants for validation purposes. We chose 1960 as a cutoff to maximize sample size while reducing the influence of early collections that could have occurred long before populations became locally adapted. However, we also examined temporal changes in phenology, as described in the next section.

Herbarium specimens were binned into geographic populations at intervals of 1° latitude for comparison with geographic populations sampled in the reciprocal transplant experiment. For each of the four 'gardens' (i.e. VCG, TIM, KSR, BEF) we calculated bootstrap estimates (1000 iterations each, with replacement) of latitudinal clines and 95% confidence intervals. In each iteration, we calculated the correlation between the population mean z-score and latitude, resampling either the phenological development time (ψ) of herbarium specimens (N = 449) binned into nine geographic 'populations' (VCG) or average flowering times of seed families sampled from six populations (N = 82).

Population Age and Cline Evolution

After validating the VCG analysis for eastern North America, we tested for parallel latitudinal clines predicted by a selection-constraint model of local adaptation to season length (18, 20, 21). This model predicts an evolutionary shift from slow to fast phenology under shorter season lengths. We also tested the CEM and PEM predictions (Fig. 1) by looking at temporal changes in the rate of cline evolution. In addition to examining clines in season length in a combined statistical model, we separately analyzed three regions with different invasion times that are also separated by natural gaps in the distribution (Fig. 2): East Coast (< 76 °W), Midwest (76 °W-101 °W), and West Coast (> 101 °W).

To examine changes in the rate of evolution over time, we first had to estimate population age, which we did using Kriging interpolation in R (52) (Fig. S1). In addition to collection dates of available herbarium records, we also included dates of other field observations available from the Global Biodiversity Information Facility (GBIF), Biodiversity Information Serving Our Nation (BISON), and Vascular Plants of Canada (VASCAN) (45, 53, 54). The resulting 11,735 records were pooled into grid units of one degree latitude and longitude, and within each grid unit, the year of the earliest observation was used to estimate the date of colonization. Our estimate of population age for each specimen was calculated by subtracting the collection year of each specimen from the locally interpolated (i.e. Kriged) year of invasion.

Statistical Methods

We modeled the observed phenological stage of each herbarium specimen (φ_i) using a logistic model (Eq. 6 in (55)), and a more detailed nonlinear least squares (NLS) regression equation

based on the Unified Gompertz curve (Eq. 8 in(55)). We report the Gompertz model here but the logistic model produced similar results (see Supplementary Methods). The analysis is based on a sigmoidal curve of the form:

[Eq. 2] $$\varphi_i = \beta_a^{\exp(-eF)}$$

where β_a is an estimate of the y-intercept and *F* is a statistical model that differs for each of three analyses, as outlined below.

First, we calculated a standardized metric of phenology development time for each herbarium specimen (ψ_i). This metric enables comparison with genetic differences observed in common garden experiments, and it can be used to visualize genetic changes over space and time. To calculate ψ_i, we modeled the observed phenological stage (φ_i) as a function of growing environment (GDD_C) using a variant of Equation 2 in which

[Eq. 3] $$F = \beta_b GDD_{c,i}$$

Predictions from this model are the predicted contribution of growing environment to phenology (E_i) shown in the VCG analysis described in Equation 1. Subtracting the observed phenological stage from the stage expected for a given GDD_c yields an estimate of the phenological development time (ψ_i). This is mathematically equivalent to taking the residuals of Equations 1 & 3 and back-transforming to the scale of φ_i.

In a second statistical model, we simultaneously estimated the relative effects of growing environment (GDD_C), season length (*SL*) and population age (*Age*). We also examined how the effects of GDD_C and *SL* change with population age of each herbarium specimen:

[Eq. 4] $$F = (\beta_b + \beta_c Age_i + \beta_d SL_i + \beta_e Age_i \times SL_i) GDD_{c,i}$$

We refer to β_b as the temperature accumulation coefficient, β_c as the population age coefficient and β_d as the season length coefficient. The coefficient β_e enables a statistical test of the hypothesis that clines strengthen over time, while β_c tests the hypothesis that cues evolve to better predict growing conditions.

To avoid problems with spatial heterogeneity in the collections (e.g. pseudoreplication due to geographical sampling bias), we calculated bootstrap means and 95% Confidence Intervals by randomly resampling, with replacement, a single specimen within each cell of a 0.1° latitude by 0.1° longitude grid (N = 1,000 iterations).

Our third statistical model was applied to specimens binned into six age classes: < 30 years, 31-60, 61-90, 91-120, 121-150, and > 151. Within each age class, we used the 0.1° resampling bootstrap model as above. Without population age as a predictor, the equation for each bin simplifies to:

[Eq. 5] $$F = (\beta_{b,t} + \beta_{d,t} SL_i) GDD_{c,i}$$

To test the CEM and PEM predictions, we fit linear and quadratic models to the bootstrapped estimates of $\beta_{b,t}$ and $\beta_{d,t}$ in Equation 5, using the median age of each time bin (*t*) as the predictor variable and the model coefficients as a response variables.

Acknowledgments

We thank SCH Barrett, LH Rieseberg, CA Rushworth, KG Turner, an anonymous handling editor and three anonymous reviewers for manuscript feedback, and the following people for generously digitizing herbarium records for our study: M Pace & B Thiers (NY Botanical Garden), D Brandon (U Memphis), CA McCormick (UNC Chapel Hill), J Shepard & T Barry (UC Davis), L Struwe (Chrysler Herbarium), R Kleinman (Zimmerman Herbarium SNM), M Nazaire (Rancho Santa Ana Botanic Garden), A Sanders (UC Riverside), P Wolf (Utah State U), D Stover (Kent State U), D Fabijan (U Alberta), S Perkins (San Juan College), K Damboise (Herbier Louis-Marie), N Cowden (Ramsey-Freer Herbarium), S Fuentes-Soriano (New Mexico State U), W Erica (Royal Museum of British Columbia), D Giblin (U Washington), M Hooker (Marion Ownbey Herbarium), A Liston (Oregon State U), A Lopez Villalobos (Fowler Herbarium). This research was supported by NSERC Discovery grant to RIC. Queen's University is situated on traditional lands of the Anishinaabe and Haudenosaunee and the authors are grateful to conduct our research from these lands.

References

1. J. D. Olden, N. LeRoy Poff, M. R. Douglas, M. E. Douglas, K. D. Fausch, Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution* **19**, 18–24 (2004).
2. R. I. Colautti, J. M. Alexander, K. M. Dlugosch, S. R. Keller, S. E. Sultan, Invasions and extinctions through the looking glass of evolutionary ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160031 (2017).
3. G. Bell, A. Gonzalez, Evolutionary rescue can prevent extinction following environmental change. *Ecology Letters* **12**, 942–948 (2009).
4. T. A. Perkins, B. L. Phillips, M. L. Baskett, A. Hastings, Evolution of dispersal and life history interact to drive accelerating spread of an invasive species. *Ecol. Lett.* **16**, 1079–1087 (2013).
5. J. M. Alexander, D. Z. Atwater, R. I. Colautti, A. L. Hargreaves, Effects of species interactions on the potential for evolution at species' range limits. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**, 20210020 (2022).
6. A. P. Hendry, M. T. Kinnison, Perspective: The pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**, 1637–1653 (1999).
7. J. G. Kingsolver, S. E. Diamond, A. M. Siepielski, S. M. Carlson, Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology* **26**, 1101–1118 (2012).
8. R. I. Colautti, J. A. Lau, Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24**, 1999–2017 (2015).
9. A. P. Hendry, *Eco-evolutionary Dynamics* (Princeton University Press, 2020).
10. O. Langlet, Two Hundred Years Genecology. *Taxon* **20**, 653–721 (1971).
11. O. Savolainen, T. Pyhäjärvi, T. Knürr, Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics* **38**, 595–619 (2007).
12. R. I. Colautti, J. L. Maron, S. C. H. Barrett, Common garden comparisons of native and introduced plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications* **2**, 187–199 (2009).
13. J. Hereford, A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist* **173**, 579–588 (2009).
14. A. M. O. Oduor, R. Leimu, M. van Kleunen, Invasive plant species are locally adapted just as frequently and at least as strongly as native plant species. *Journal of Ecology* **104**, 957–968 (2016).
15. A. K. Knapp, *et al.*, Past, present, and future roles of long-term experiments in the LTER network. *BioScience* **62**, 377–389 (2012).
16. P. G. Grant, R. G. Grant, *40 Years of Evolution: Darwin's Finches on Daphne Major Island* (Princeton University Press, 2014).

17. J. L. Montague, S. C. H. Barrett, C. G. Eckert, Re-establishment of clinal variation in flowering time among introduced populations of purple loosestrife (*Lythrum salicaria*, Lythraceae). *J Evol Biol* **21**, 234–245 (2008).
18. R. I. Colautti, S. C. H. Barrett, Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science* **342**, 364–366 (2013).
19. R. I. Colautti, S. C. H. Barrett, Natural selection and genetic constraints on flowering phenology in an invasive plant. *Int. J. Plant Sci.* **171**, 960–971 (2010).
20. R. I. Colautti, S. C. H. Barrett, Population divergence along lines of genetic variance and covariance in the invasive plant *Lythrum salicaria* in eastern North America. *Evolution* **65**, 2514–2529 (2011).
21. R. I. Colautti, C. G. Eckert, S. C. H. Barrett, Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1799–1806 (2010).
22. S. J. Gould, N. Eldredge, Punctuated equilibrium comes of age. *Nature* **366**, 223–227 (1993).
23. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018).
24. S. J. Franks, S. Sim, A. E. Weis, Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *PNAS* **104**, 1278–1282 (2007).
25. J. R. Etterson, *et al.*, Project Baseline: An unprecedented resource to study plant evolution across space and time. *American Journal of Botany* **103**, 164–173 (2016).
26. S. J. Franks, E. Hamann, A. E. Weis, Using the resurrection approach to understand contemporary evolution in changing environments. *Evol Appl* **11**, 17–28 (2017).
27. L. J. Weider, P. D. Jeyasingh, D. Frisch, Evolutionary aspects of resurrection ecology: Progress, scope, and applications—An overview. *Evol Appl* **11**, 3–10 (2018).
28. P. L. M. Lang, F. M. Willems, J. F. Scheepens, H. A. Burbano, O. Bossdorf, Using herbaria to study global environmental change. **221**, 110–122 (2019).
29. C. G. Willis, *et al.*, Old Plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology & Evolution* **32**, 531–546 (2017).
30. B. P. Hedrick, *et al.*, Digitization and the future of natural history collections. *BioScience* **70**, 243–251 (2020).
31. C. C. Davis, C. G. Willis, B. Connolly, C. Kelly, A. M. Ellison, Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. *American Journal of Botany* **102**, 1599–1609 (2015).
32. J. Merilä, A. P. Hendry, Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications* **7**, 1–14 (2014).
33. D. Q. Thompson, R. L. Stuckey, E. B. Thompson, *Spread, impact, and control of purple loosestrife (*Lythrum salicaria*) in North American wetlands* (U.S. Fish and Wildlife Service, 1987).
34. J. L. Williams, B. E. Kendall, J. M. Levine, Rapid evolution accelerates plant population spread in fragmented experimental landscapes. *Science* **353**, 482–485 (2016).
35. M. Szűcs, *et al.*, Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *PNAS* **114**, 13501–13506 (2017).
36. S. Peischl, I. Dupanloup, M. Kirkpatrick, L. Excoffier, On the accumulation of deleterious mutations during range expansions. *Molecular Ecology* **22**, 5972–5982 (2013).
37. J. B. Losos, Convergence, Adaptation, and Constraint. *Evolution* **65**, 1827–1840 (2011).
38. K. M. Lee, G. Coop, Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180236 (2019).
39. J. Arendt, D. Reznick, Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution* **23**, 26–32 (2008).
40. Y. J. Chun, J. D. Nason, K. A. Moloney, Comparison of quantitative and molecular genetic variation of native vs. invasive populations of purple loosestrife (*Lythrum salicaria* L., Lythraceae). *Molecular Ecology* **18**, 3020–3035 (2009).
41. K. Olsson, J. Ågren, Latitudinal population differentiation in phenology, life history and flower morphology in the perennial herb *Lythrum salicaria*. *J. Evol. Biol.* **15**, 983–996 (2002).
42. C. Liu, C. Wolter, W. Xian, J. M. Jeschke, Most invasive species largely conserve their climatic niche. *PNAS* **117**, 23643–23651 (2020).
43. J. Novembre, N. H. Barton, Tread lightly interpreting polygenic tests of selection. *Genetics* **208**, 1351–1355 (2018).
44. K. E. Lohmueller, The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet* **10** (2014).
45. P. Desmet, L. Brouillet, Database of Vascular Plants of Canada (VASCAN): a community contributed taxonomic checklist of all vascular plants of Canada, Saint Pierre and Miquelon, and Greenland. *PhytoKeys*, 55–67 (2013).

46. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
47. S. R. A. Shamsi, F. H. Whitehead, Comparative eco-physiology of *Epilobium hirsutum* L. and *Lythrum salicaria* L.: II. Growth and development in relation to light. *Journal of Ecology* **62**, 631–645 (1974).
48. Y. J. Chun, Phenotypic plasticity of introduced versus native purple loosestrife: univariate and multivariate reaction norm approaches. *Biol Invasions* **13**, 819–829 (2011).
49. C. J. Lindgren, D. Walker, Growth rate, seed production, and assessing the spatial risk of *Lythrum salicaria* using growing degree-days. *Wetlands* **32**, 885–893 (2012).
50. M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, T. G. Houston, An overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric & Oceanic Technology* **29**, 897–910 (2012).
51. E. Ferrarese, R. J. Garono, S. Schooler, "Assessment of Purple Loosestrife Biocontrol Agent Populations on the Columbia River" (Earth Design Consultants, Inc., 2009).
52. B. Gräler, E. Pebesma, G. Heuvelink, Spatio-temporal interpolation using gstat. *The R Journal* **8**, 204–218 (2016).
53. GBIF.org, GBIF: The Global Biodiversity Information Facility (2017).
54. US Geological Survey, "Species occurrence data for the Nation—USGS Biodiversity Information Serving Our Nation (BISON)" (U.S. Geological Survey, 2019).
55. K. M. C. Tjørve, E. Tjørve, A proposed family of unified models for sigmoidal growth. *Ecological Modelling* **359**, 117–127 (2017).

Figures and Tables

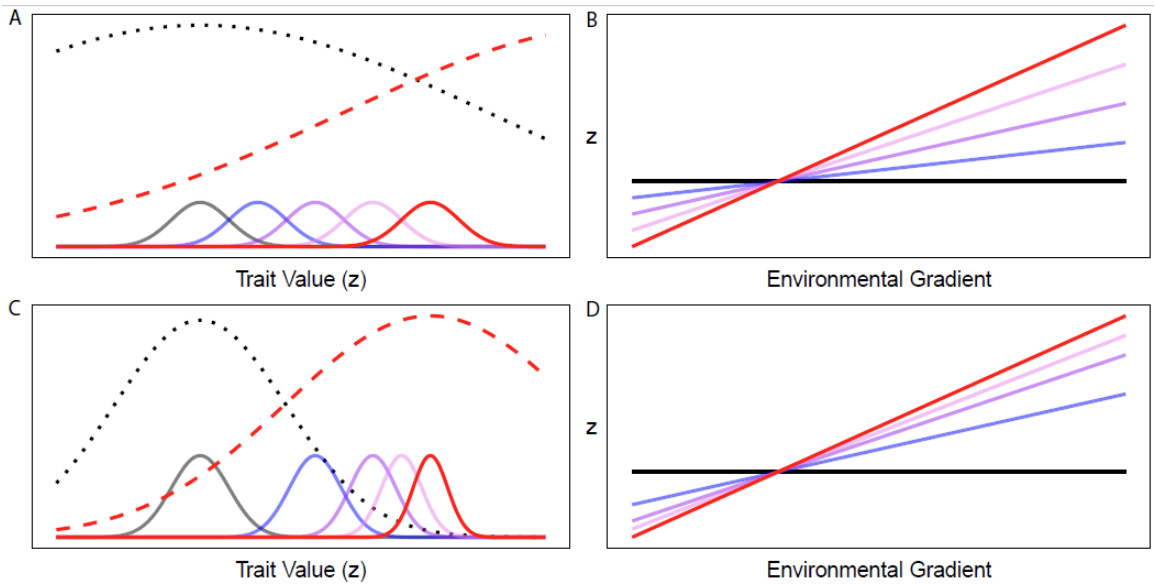

Figure 1. Models of adaptive evolution in a novel environment of a single population (A & C) and multiple populations along an environmental gradient (B & D). In the single population model (A & C), the distribution of phenotypes (*z*) in ancestral populations (solid grey curves) is maintained by stabilizing selection in the ancestral environment (dotted grey curve). Colonization of a novel environment (or a change in local environment) shifts the fitness surface to a new optimum (red dashed curve). A trait under selection evolves towards the new optimum over time (blue through red curves). When adaptive evolution occurs in many populations located along an environmental gradient (B & D), then clines (i.e. slopes) increase in magnitude over time. The Continuous Evolution Model (CEM; A & B) predicts a relatively constant rate of evolution through time. This could occur when (i) stabilizing selection is weak relative to the difference in optima, resulting in approximately linear selection differentials (e.g. dashed red line in A), and (ii) genetic variation introduced by mutation and migration is relatively constant through time. The Punctuated Evolution Model (PEM; C & D) predicts a decelerating rate of evolution. This could occur when stabilizing selection is strong relative to the difference in optima (e.g. dashed red line in C), or when standing genetic variation decreases through time for the trait under selection.

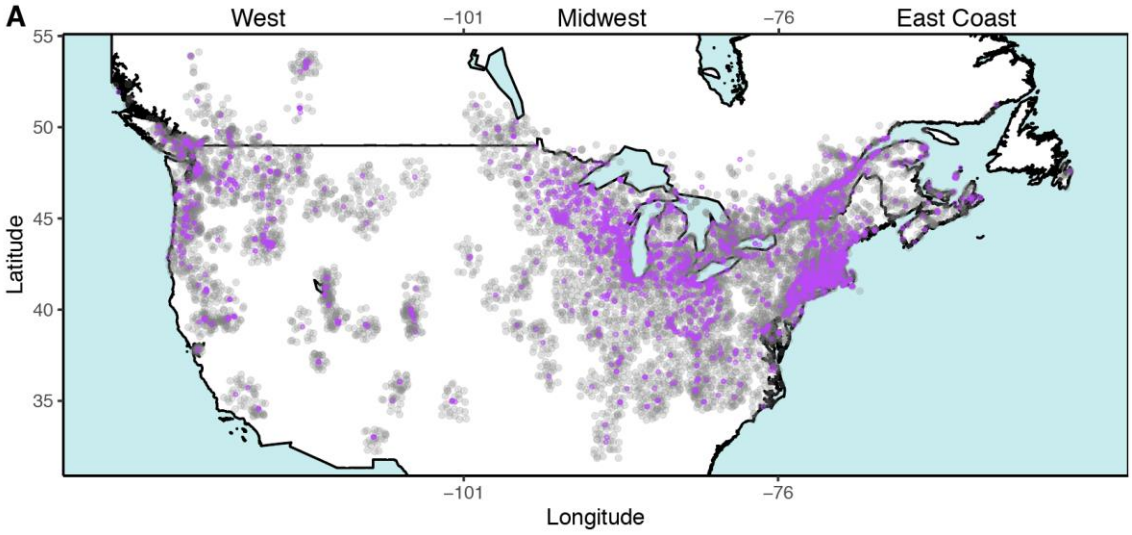

Figure 2. Map of the United States and Canada showing locations of 3,429 *Lythrum salicaria* specimens (purple dots) collected between 1866 and 2016, and locations of 6,303 weather stations (light grey dots) used to interpolate local climate data. Specimens are classified into three geographical regions with different invasion histories: West (< -101° longitude), Midwest (-76° to -101°), and East Coast (> -76° longitude).

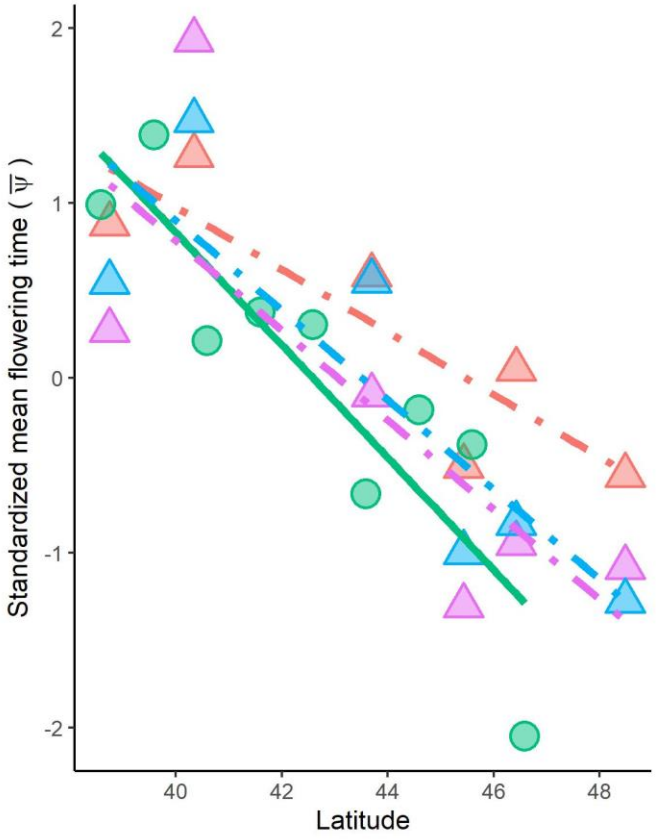

Figure 3. Validation of clines observed in the Virtual Common Garden (VCG). Standardized estimates (z-score) of mean phenological development time ($\bar{\psi}$) were inferred from contemporary herbarium specimens from eastern North America, binned by one degree of latitude (green circles). This cline is similar to average z-scores of flowering times estimated for six populations grown at each of three reciprocal transplant sites spanning 10° of latitude: Timmins, Ontario (blue triangles), Koffler Scientific Reserve near Newmarket, Ontario (purple triangles), and Blandy Experimental Farm near Boyce, Virginia (red triangles).

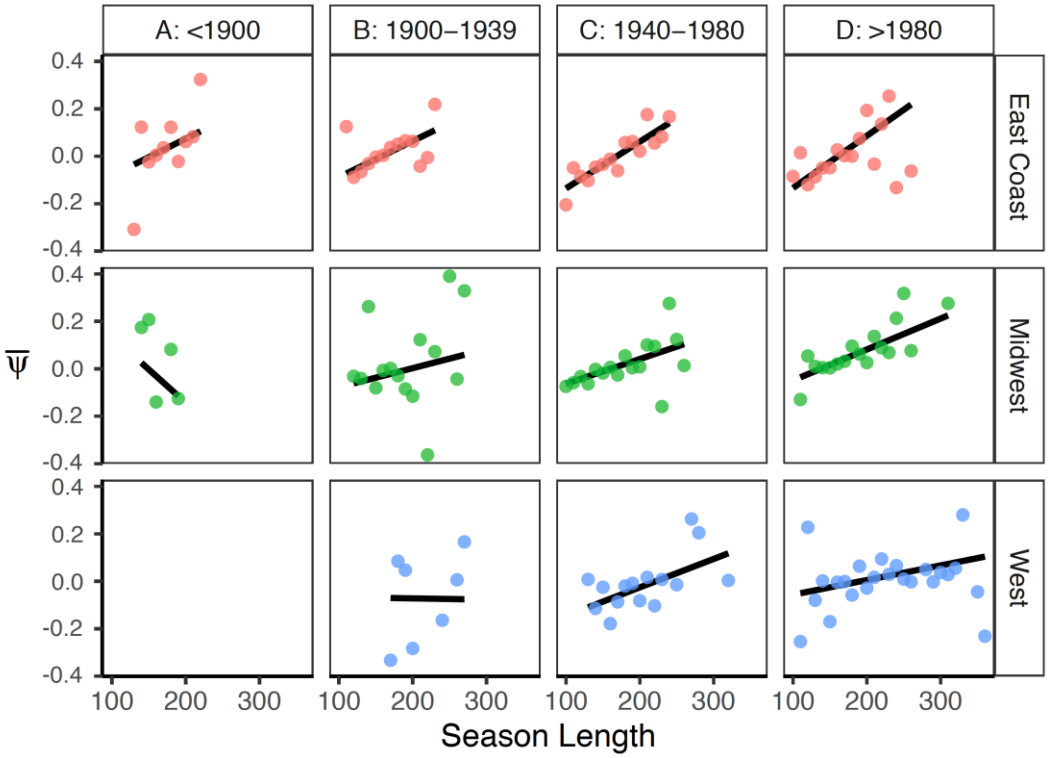

Figure 4. Spatial (rows) and temporal variation (columns) in phenological clines of *Lythrum salicaria* as it spread across North America. Clines in the East Coast (red dots, top row), Midwest (green dots, middle row), and West regions (blue dots, bottom row) are divided by time of collection, binned into four eras (columns). Each bivariate plot shows the relationship between the mean phenological development time of herbarium specimens ($\overline{\psi}$) and the mean estimated season length, averaged across specimens binned every ten days of season length.

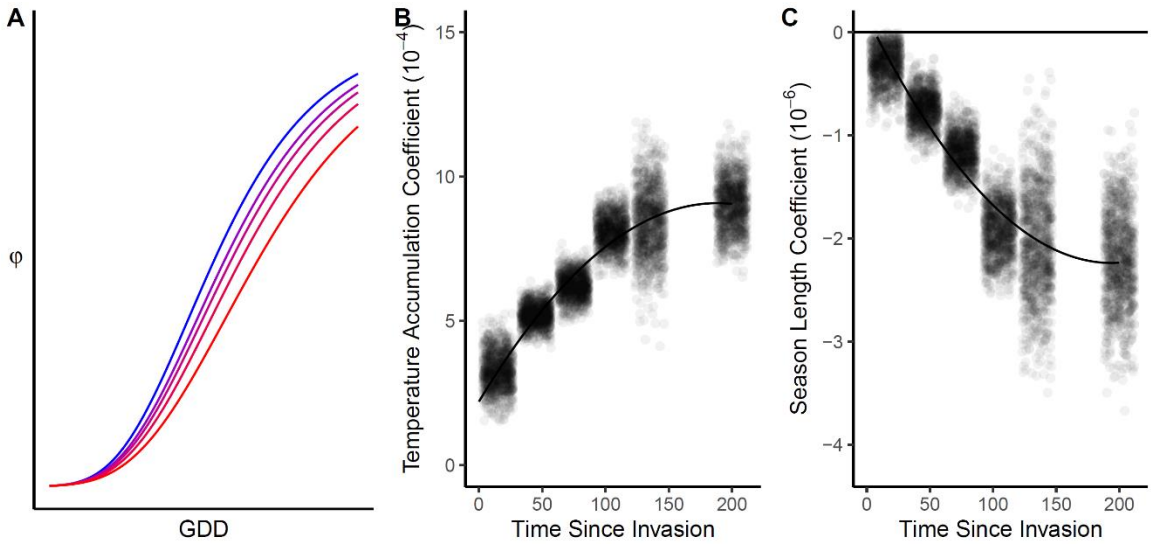

Figure 5. A nonlinear least-squares (NLS) model testing temporal changes in phenology clines. (A) A model of observed phenological stage (φ) as a function of local growing conditions (growing-degree days at the time of collection; GDD_C) for five different season lengths (SL) parameters ranging from short/northern (blue) to long/southern (red). (B) Bootstrap estimates of $\beta_{b,t}$, the model coefficient for GDD_C. (C) Bootstrap estimates of $\beta_{d,t}$, the model coefficient for season length (SL). Bootstrap estimates include 1,000 spatially explicit resampling iterations (with replacement) with specimens binned into one of six classes of population age (i.e. time since establishment). Lines in B and C show quadratic linear regression of bootstrap estimates.