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

**Authors:** Yihan Wu and Robert I. Colautti*

**Affiliations:** Department of Biology, Queen's University, Kingston, Ontario, Canada

*Correspondence to: robert.colautti@queensu.ca

# 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. While field experiments have demonstrated that adaptive evolution can occur quickly, our understanding of the duration and geographical extent of contemporary evolution in natural systems remains limited. 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 estimate genetic differences among phenotypes observed in natural history collections. Reconstructing 150 years of evolution in *Lythrum salicaria* (purple loosestrife) as it invaded across 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 evolves rapidly and repeatedly along parallel climatic gradients during the first century of evolution. However, the rate of microevolution stalls thereafter, recapitulating macroevolutionary stasis observed in the fossil record. Our study demonstrates why preserved specimens are a critical resource for understanding limits to evolution in natural. Our results show predictability of evolution emerging at a continental scale across 15 decades of rapid, adaptive evolution.


## Significance

Adaptive evolution can help species to persist in new environments. The fossil record contains many examples of phenotypic stasis punctuated by rapid evolution, with distinct lineages converging on similar phenotypes over geological timescales. In contrast, the spatio-temporal dynamics of evolution over ecological timescales are largely unknown. Here, we use a computational approach to reconstruct 15 decades of evolution in an invasive plant as it spread across North America. Flowering phenology evolves in parallel throughout the range but stalls after about a century. This punctuated, convergent evolution recapitulates long-term dynamics in the fossil record, suggesting constraints on adaptation that are not evident for the first hundred years.

# Introduction

Global biodiversity in the Anthropocene is threatened by the jointly homogenizing effects of population extirpation and biological invasion (Olden *et al.* 2004). These outcomes lie along a spectrum of population growth trajectories that depend fundamentally on environmental (mal)adaptation (Colautti *et al.* 2017). Adaptive evolution in novel and changing environments can rescue populations from extinction (Bell & Gonzalez 2009) and facilitate the spread of invasive species (Perkins *et al.* 2013). However, genetic constraints limit the evolutionary response to selection in natural populations such that the balance of adaptation and constraint within a species partly determine its ecological niche (Blows & Hoffmann 2005). Species distribution models and management strategies under current and future global change would therefore benefit from a better understanding of how natural selection and genetic constraints affect the rate, duration, and geographical extent of adaptive evolution in natural populations.

Selection gradients and rates of phenotypic evolution have been studied extensively in native populations for a variety of taxa (Hendry & Kinnison 1999; Kingsolver *et al.* 2012; Colautti & Lau 2015; Hendry 2020). In contrast, little is known about selection operating in natural populations, due to an apparent researcher bias towards studying 'pristine' natural systems (Colautti & Lau 2015). Moreover, the duration and geographical extent of adaptive evolution are difficult to assess using conventional methods, particularly when one is interested in variation throughout a species range over decades to centuries timescales. Common garden studies involving genotypes from contemporary populations can offer insight into evolutionary changes in response to local environments, but provide only a snapshot of evolutionary change (Langlet 1971; Savolainen *et al.* 2007; Colautti *et al.* 2009; Hereford 2009; Oduor *et al.* 2016). 'Long-term' ecological studies reveal 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 (Knapp *et al.* 2012; Grant & Grant 2014). Therefore, it remains unclear how the rate and duration of adaptive evolution varies throughout species' ranges, and whether evolution can be sustained over the decade to century timescales relevant to managing biodiversity for the 21$^{st}$ century.

Invasive species provide opportunities to study evolution in action, but inferences from contemporary populations are complicated by interactions between demography (i.e. invasion history) and spatio-temporal variation in natural selection (Colautti & Lau 2015). Clines in growth and phenology have been identified in many invasive plants, but the rate, duration, and adaptive significance of cline evolution remains uncertain for most of these species (Colautti *et al.* 2009; Colautti & Lau 2015). For example, 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 (Montague *et al.* 2008; Colautti & Barrett 2013). Local adaptation in these populations likely evolved as a consequence of both (i) trade-offs between growth and phenology, and (ii) a latitudinal cline in the strength of directional selection on flowering time (Colautti & Barrett 2010, 2011, 2013; Colautti *et al.* 2010). However, this research relies on inferences from contemporary genotypes within a small portion of the North American distribution. It is therefore uncertain whether parallel clines exist across North America, and how quickly these adaptive clines evolved during invasion.

Generally, 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) would predict that the strength and direction of natural selection are relatively consistent through time as populations evolve toward their locally adaptive fitness optima (Fig. 1a,b). Alternatively, a punctuated equilibrium model (PEM) would predict an early burst of evolution, followed by a static equilibrium stage with little net directional change (Gould & Eldredge 1993). This would occur as populations approach their fitness optima and/or lack sufficient genetic variation to maintain an evolutionary response (Fig. 1c,d). 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 for any study species requires simultaneously reconstructing evolutionary change at multiple time points and at many locations throughout the range.

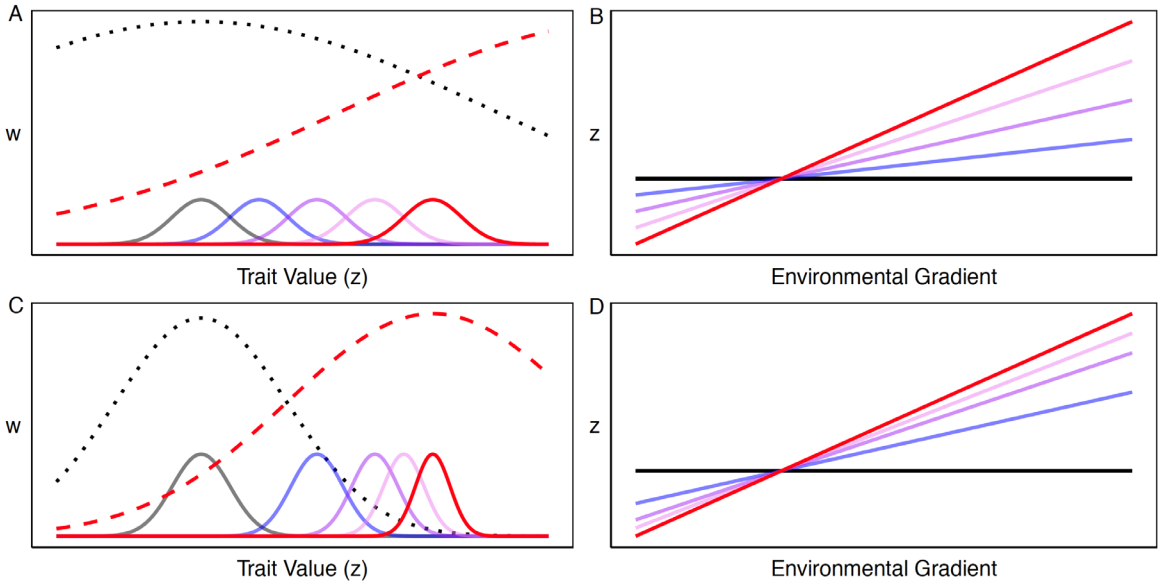



**Figure 1.** Models of adaptive evolution in a novel environment of a single population (A & C) and multiple populations along a geographical gradient (B & D). In the single population model (A & C), ancestral populations (solid grey curves) are adapted to their ancestral environment (dotted grey curve). After colonization of a novel environment (or a change in local environment), a trait under selection evolves over time (blue through red curves) toward a new fitness optimum (dashed red curve). When adaptive evolution occurs in populations located along an environmental gradient (B & D), then clines in a trait (z) under selection increase in magnitude over time. The Continuous Evolution Model (CEM; A & B) predicts a relatively constant rate of evolution through time. The Punctuated Evolution Model (PEM; C & D) predicts a deceleration in the rate of evolution.

Temporal sampling of genetic changes in natural populations provides some of the strongest evidence for adaptive evolution in nature (Franks *et al.* 2007; Grant & Grant 2014), but these experiments are difficult to replicate over large spatial and temporal scales (Etterson *et al.* 2016; Franks *et al.* 2017; Weider *et al.* 2018). In contrast, millions of preserved specimens maintained in global natural history collections provide a detailed record of phenotypic change over decades to centuries past, holding potential for reconstructing evolutionary trajectories (Willis *et al.* 2017; Lang *et al.* 2019). However, evolutionary shifts in phenotypic traits observed under natural field conditions are difficult to distinguish from developmental plasticity associated with different environments (Merilä & Hendry 2014). Thus, it would be valuable to incorporate data from controlled experiments to empirically model the effects of development environment and genetic differentiation on phenotypes observed in the field. Such a model could enable tests of adaptive evolution and genetic constraint across broad temporal and spatial scales.

Here, we introduce the virtual common garden (VCG) as a novel computational approach, parameterized by field experiments, to model phenotypic plasticity and genetic differentiation in phenotypic data from historical archives. As such, the VCG can be thought of as a new form of observational experiment for eco-evolutionary inference. We use this approach to reconstruct 150 years of phenotypic evolution in the 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 in the form of parallel phenology clines replicated across North America; (iii) to measure changes in the rate of cline evolution through time in order to test the CEM and PEM models shown in Figure 1.

# Methods

### *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*)(Desmet & Brouillet 2013) and v) correspondence with 19 university herbaria to obtain more images (see acknowledgements in main text). 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 (Schneider *et al.*

2012). We did this for the primary meristem, unless it was clearly damaged in which case the longest inflorescence was measured. Unpollinated flowers in *L. salicaria* senesce and fall off of the stem but leave a distinctive scar; these were included as fruits as they represent a post-flowering phenology. Using these measurements, we calculated a scale-free phenological index ($\varphi$) for each herbarium specimen $i$:

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

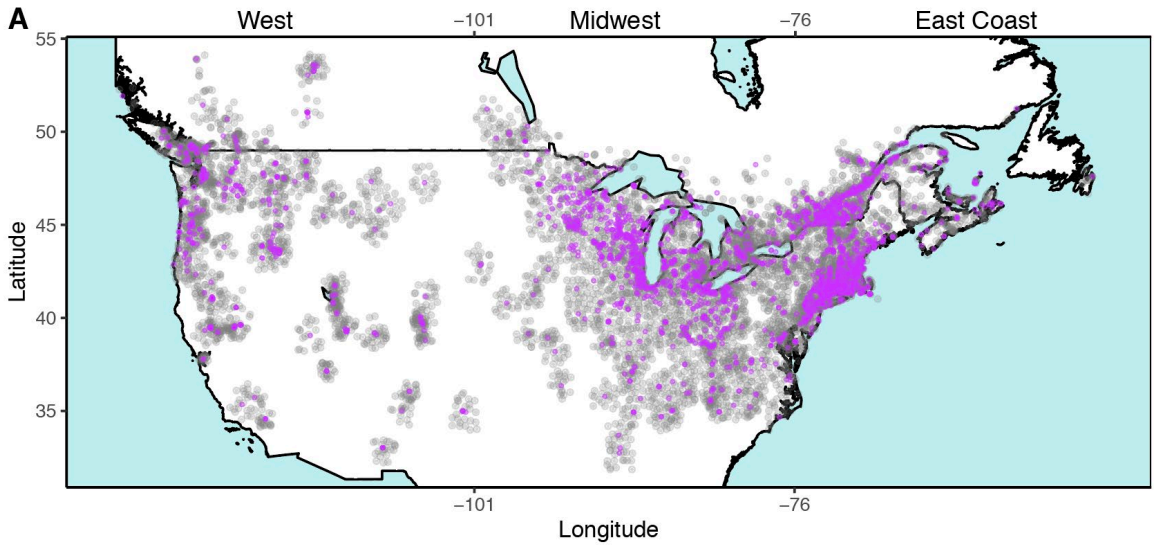
Values of $\varphi$ range from 0 (early) to 1 (late) as a measure of phenological stage at the time of collection. Thus, on any given collection date, specimens with a phenology index closer to 0 represent phenotypes early in their phenology whereas those with an index closer to 1 represent phenotypes 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 (Davis *et al.* 2015). Spatial and temporal clustering of samples is also likely when multiple specimens are often collected in sampling excusions 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 (Shamsi & Whitehead 1974; Chun 2011; Lindgren & Walker 2012; Colautti & Barrett 2013). Both *SL* and $GDD_C$ were interpolated from computation of daily weather data available from the Global Historical Climatology Network-Daily database (Menne *et al.* 2012). The algorithms and code for the computational interpolation are available on the Dryad Database (DOI: Reviewer Note: zip files for Dryad are uploaded with manuscript and also available in public Repo on GitHub https://github.com/ColauttiLab/WuColauttiHerbarium). Briefly, we collected temperature data for each herbarium specimen from up to 20 of the closest weather stations located within 0.5º 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 (Shamsi & Whitehead 1974; Montague *et al.* 2008; Ferrarese *et al.* 2009; Lindgren & Walker 2012). A length of 10 days at the start of the season was chosen to avoid shorter intervals of abnormally warm winter or spring

190 temperatures where growth would be unlikely. We then used the season start date to calculate the
191 cumulative growing degree days until the day of collection ($GDD_C$). In summary, $GDD_C$
192 characterizes the local growing environment in order to correct for variation in growing
193 conditions and sampling date, whereas variation in season length ($SL$) is predicted to drive
194 differentiation via local adaptation to climate.




195

196 **Figure 2.** Map of the United States and Canada showing locations of 3,429 *Lythrum*
197 *salicaria* specimens (purple dots) collected between 1866 and 2016, and locations of
198 6,303 weather stations (light grey dots) used to interpolate local climate data.
199 Specimens are classified into three geographical regions with different invasion
200 histories: West (< -101° longitude), Midwest (-76° to -101°), and East Coast (> -76°
201 longitude).

202

203 ***Virtual Common Garden (VCG)***

204 Field surveys and common garden experiments with *L. salicaria* have demonstrated both
205 plastic and genetic variation for flowering phenology – warmer temperatures accelerate
206 phenological development, whereas natural selection has caused the evolution of earlier
207 flowering in short growing seasons (Shamsi & Whitehead 1974; Colautti & Barrett 2013). By
208 comparison, population-by-environment interactions are relatively weak, even across field sites
209 spanning a latitudinal gradient of 1,000 km (Colautti & Barrett 2013). Leveraging these
210 observations, the VCG uses temperature accumulation data to model phenology as the additive
211 effects of genotype and plasticity. Since we cannot retroactively grow herbarium specimens in a
212 common garden, we instead correct the observed phenological stage of each specimen ($\varphi_i$) for
213 variation in the local growing environment ($E_i$) to estimate a relative measure of phenological
214 development time due to genetic factors ($\psi_i$):

215 $$\psi_i = E_i - \varphi_i$$

As described later, $E_i$ is estimated from a nonlinear least-squares model of phenology as a logistic function of $GDD_C$. The equation for phenological development time ($\psi$) 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 ($\varphi$) is calculated directly from inflorescence measurements. The phenological development time ($\psi$) is an estimate of relative phenology after controlling for local growing environment. Phenological stage ($\varphi$) 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 ($\psi$) ranges from -1 (fast phenology) to 1 (slow phenology) and predicts genetic effects on phenology. As such, $\psi$ 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 VGC, we compare the predicted phenological development time ($\psi_i$) 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 10º 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 (Colautti & Barrett 2013).

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. This date was chosen 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 1º latitude for comparison with geographic populations sampled in the reciprocal transplant experiment. We used four separate bootstrap models, (1000 iterations each, with replacement) to estimate latitudinal clines and 95% confidence intervals for each of the four 'gardens' (i.e. VCG, TIM, KSR, BEF). In each iteration, we calculated the correlation between the population mean z-score and latitude, resampling either the phenological index ($\psi$) 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 (Colautti *et al.* 2010; Colautti & Barrett 2011, 2013). This model predicts an evolutionary shift from slow to fast phenology under shorter season lengths. We also tested the CEM and GEM 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 (Gräler *et al.* 2016). 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) (Desmet & Brouillet 2013; GBIF.org 2017; US Geological Survey 2019). Data 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 Test of Phenology Evolution*

We modeled phenological development using a nonlinear least squares (NLS) regression equation based on the Universal Gompertz curve (Tjørve & Tjørve 2017), which is a sigmoidal curve of the form:

$$\varphi = \beta_a^{\,e^{-eF}}$$

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

To calculate the standardized phenology development time of each specimen ($\psi_i$), we modeled the phenological index ($\varphi$) as a function of growing environment ($GDD_C$):

$$F = \beta_b GDD_c$$

Predictions from this model are the environmental effects on phenology ($E_i$) in the VCG analysis described earlier.

To test for evidence of evolution, 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:

$$F = (\beta_b + \beta_c Age + \beta_d SL + \beta_e Age \cdot SL) GDD_c$$

We refer to $\beta_b$ as the temperature accumulation coefficient, $\beta_c$ as the population age coefficient and $\beta_d$ as the season length coefficient. The coefficient $\beta_e$ estimates the interactive effect of age-by-season length on phenological development. To avoid problems with spatial heterogeneity in the collections (e.g. geographical sampling bias), we fit and tested the model with 1000 bootstrap datasets created by randomly sampling a single specimen within each cell of a 0.1° latitude by 0.1° longitude grid.

In addition to a single model testing evolution on the full set of herbarium specimens, we also performed individual models for specimens binned into six age classes based on our Kriging model: < 30 years, 31-60, 61-90, 91-120, 121-150, and > 151. We used a similar bootstrap model as described above to account for spatial heterogeneity. Without population age as a predictor, the equation for each bin simplifies to:

$$F = (\beta_b + \beta_d \cdot SL) \cdot GDD_c$$

To visualize how model coefficients changed over time, we fit linear and quadratic models to the bootstrapped estimates of $\beta_b$ and $\beta_d$, using the age of each time bin as the predictor variable, and each model coefficient as the response variable.

## Results

### *Herbarium Specimens*

Our spatially-weighted interpolation of Growing Degree Days ($GDD_C$) and Season Length ($SL$) for each of the 3,429 digitized herbarium specimens was based on >12 million temperature records, used to reconstruct 62,208 annual temperature accumulation curves from 6,303 weather stations (Fig. 2).

A final set of measurements from all digitized herbarium specimens (Fig. 2) that met inclusion criteria are available online in the Dryad Database (DOI: Reviewer Note: zip files for Dryad are uploaded with manuscript and also available in public Repo on GitHub https://github.com/ColauttiLab/WuColauttiHerbarium) 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). Most 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 represented a broad range of phenological stage from a few buds to fully mature fruits. There was no clear 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 aforementioned relationship between season length and latitude (Fig. S5).

### *VCG Validation*

To validate the VCG we compared (i) latitudinal clines in development time ($\psi$) of herbarium specimens with (ii) days to first flower in three real-world common garden experiments in eastern North America (data from (Colautti & Barrett 2013)), each standardized to a mean of zero and unit variance. This comparison involves heterogenous phenology measurements (i.e. $\psi$ and days to first flower) and growing conditions (i.e. a virtual common garden and three garden sites spanning 10° of latitude), yet the estimated latitudinal clines were remarkably consistent (Fig. 3). The correlation coefficient 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).

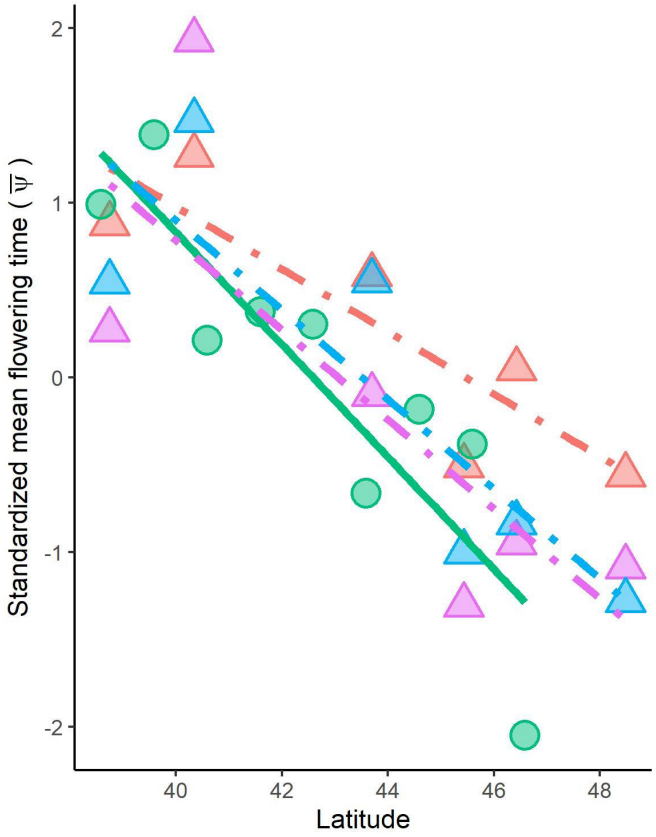



**Figure 3.** Validation of the Virtual Common Garden (VCG). Standardized estimates (z-score) of mean phenological development time ($\bar{\psi}$) were inferred from contemporary herbarium specimens from eastern North America, binned by 1 degree of latitude (green circles). This cline is similar to average 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).

### *Cline evolution*

Clines in phenology with growing season were observed in all three geographic regions for specimens collected after 1980, and generally weaken back through time (Fig. 4, Table S3). We used the NLS model as a formal statistical test of evolutionary changes using dates of establishment estimated from Kriging (Fig. S1). The NLS model relates the observed phenological stage ($\varphi$) of a specimen to the accumulated temperature it experienced up to the day of collection and the total season length in which it evolved (Fig. 5a, Table S4). The estimated model coefficients for linear (CEM) and quadratic (PEM) models are reported in Table S5 for the effect of temperature accumulation on phenology ($\beta_b$), and Table S6 shows the coefficients

for the effect of season length on phenology ($\beta_d$). The quadratic terms were significant for both $\beta_b$ (Likelihood Ratio Test: $\chi^2$ = 5448, p < 0.001) and $\beta_d$ ($\chi^2$ =3045, p < 0.001) based on likelihood Ratio, tests with a ~15% increase in $R^2$ values ($\beta_b$: 0.763 to 0.876; $\beta_d$: 0.713 to 0.810). This is shown in Figure 5 wherein the coefficients for $\beta_b$ (Fig. 5b) and $\beta_d$ (Fig. 5c) stabilize around 100 years post-establishment.

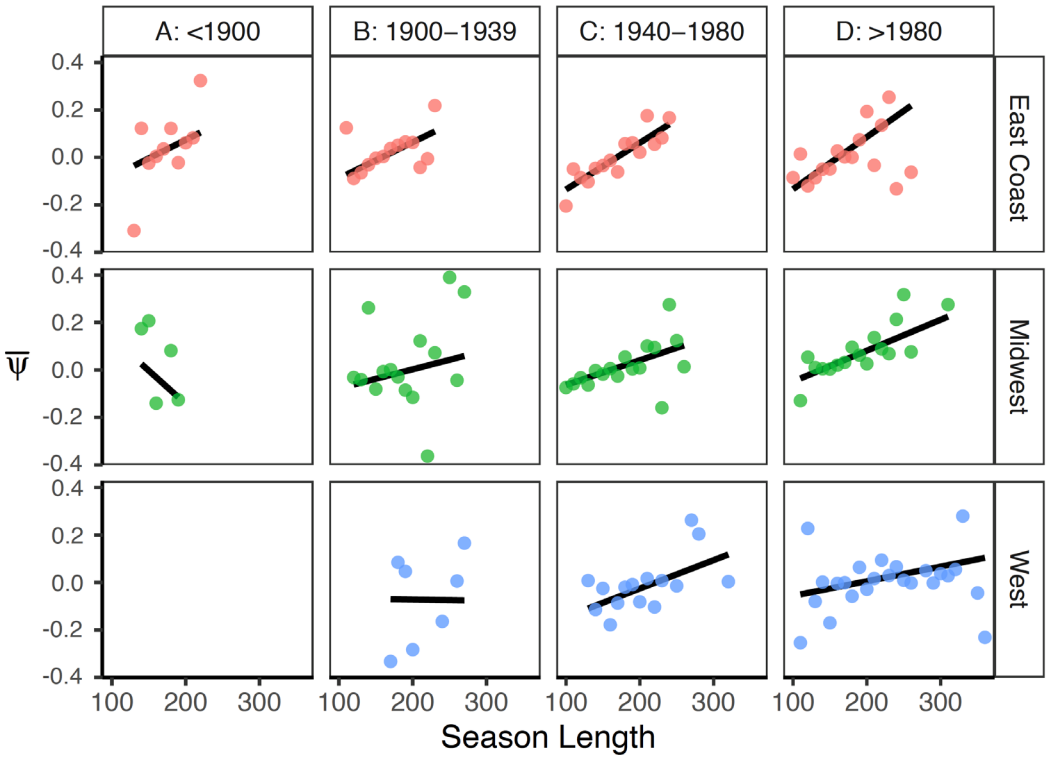



**Figure 4.** Spatial (rows) and temporal variation (columns) in phenological clines of *Lythrum salicaria* as it spread across North America. Clines in the East Coast (red dots, top row), Midwest (green dots, middle row), and West regions (blue dots, bottom row) are divided by time of collection, binned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

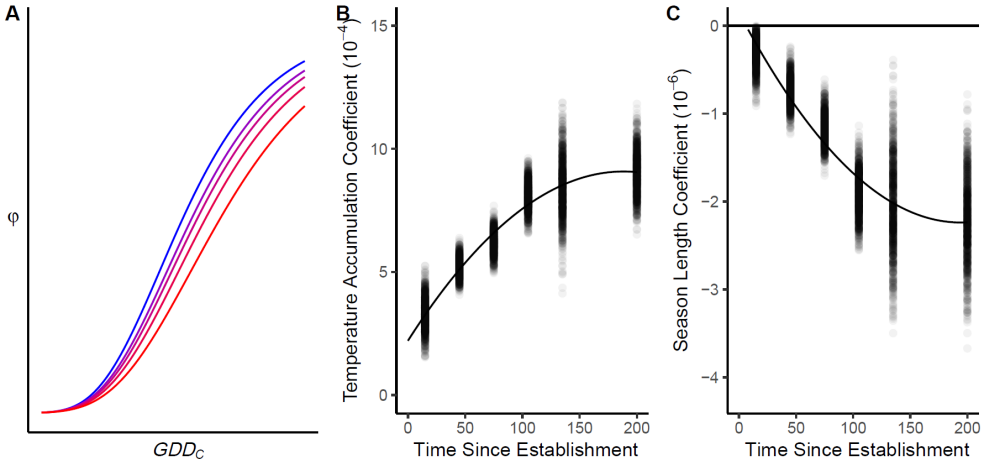



**Figure 5.** A nonlinear least-squares (NLS) model testing temporal changes in phenology clines. (A) A model of observed phenological stage ($\varphi$) as a function of local growing conditions (growing-degree days at the time of collection; $GDD_C$) for five different season lengths ($SL$) parameters ranging from short/northern (blue) to long/southern (red). (B) Bootstrap estimates of $\beta_b$, the model coefficient for $GDD_C$. (C) Bootstrap estimates of $\beta_d$, 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).

## Discussion

Geographical clines in morphological, physiological and life-history traits are common to many taxa, but the timescale and geographical extent of cline evolution has been difficult to assess. In this study, we introduced the Virtual Common Garden (VCG) analysis as a computational approach to test evolutionary hypotheses using natural history collections by accounting for variation in local growing environments. Below, we explore the relevance of our findings to evolution in novel environments and the spread of invasive species.

Feedbacks between colonization and trait evolution can facilitate rapid range expansion (Perkins *et al.* 2013; Williams *et al.* 2016; Szűcs *et al.* 2017), but clines per se are not evidence for adaptive evolution because non-adaptive and maladaptive clines can evolve through repeated founder events during spread (Peischl *et al.* 2013; Colautti & Lau 2015). A reciprocal transplant experiment involving *L. salicaria* demonstrated that flowering time clines are adaptive in part of eastern North America (Colautti & Barrett 2013), but it has been less clear whether natural selection is a generally a causal agent of cline formation or merely works in a supporting role to

maintain clines that evolved through stochastic processes. Molecular genetic markers can help to distinguish adaptive evolution from non-adaptive, demographic effects. However, this approach has limitations for quantitative traits like phenology, for organisms like *L. salicaria* that lack an annotated genome, and for populations with complex demographic histories like invasive species (Lohmueller 2014; Colautti & Lau 2015; Novembre & Barton 2018). In contrast, a VCG analysis of specimens sampled across a large geographical area at multiple time points can test adaptive hypotheses because genetic drift and founder effects should produce phenological clines with random slopes throughout the introduced range, whereas climate adaptation yields parallel clines.

We found that clines in phenology are indeed replicated across North America (Fig. 4), adding support for the theory that rapid, adaptive evolution of flowering time facilitated the spread of *L. salicaria* across the continent. An important assumption inherent in the VCG is that phenology can be predicted by the additive effects of genotype and environment. The parallel clines shown in Figure 3 support this assumption because the slopes would differ if genotypes had a strong statistical interaction with environmental characteristics that covary with latitude (e.g. photoperiod, temperature). Instead, we observe a constancy of clines across rearing environments, both real and virtual.

We 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 PEM over the CEM, with a constant rate of cline evolution for ~100 years followed by a contemporary period of relative stasis (Fig. 5). Indeed, the NLS model suggests two different but complementary mechanisms of cline evolution. Perhaps the more intuitive is the strengthening of the season length coefficient from near-zero (Fig. 5c). In other words, season length is a significant predictor of phenology in the model for populations established >50 years ago but this effect weakens for younger populations. This is consistent with the continent-wide evolution of flowering phenology in response to variation in season length, as predicted by a selection-constraint model developed for populations in eastern North America (Colautti *et al.* 2010; Colautti & Barrett 2013). In addition to the NLS results, the selection-constraint model and the PEM are further supported by the region-specific temporal analysis of clines (Fig.4). East Coast populations were well established by the time of first collection and show a relatively stable cline varying with season length. Contemporary populations in the Midwest have a similar cline that weakens back toward the time of establishment (~1900). In contrast, West Coast populations did not establish until the 1920s, and the correlation coefficient in contemporary populations has not yet reached the same magnitude as the two older regions (Fig. 4).

Complementary to evolution of phenology as an adaptive response to season length, the NLS model reveals a second potential mechanism of adaptive evolution via a strengthening of the temperature accumulation coefficient with time since establishment (Fig. 5b). This effect is not dependent on season length but rather shows a general strengthening through time of the relationship between the phenological index ($\varphi$) and growing degree days ($GDD_C$). This response is consistent with the evolution of environmental cues, growth rates, development

times, and other mechanisms that allow genotypes to fine-tune phenologies to maximize survival and reproduction in local climates.

Together, our analysis supports convergent evolution of local adaptation throughout North America over a timescale of ~100 years, followed by a contemporary period of relative stasis. The hundred-year response to selection may not be unique to *L. salicaria*, as many other Eurasian plants introduced to North America also exhibit genetically-based clines (Colautti *et al.* 2009) and segregate genetic variation for quantitative traits at levels comparable to native genotypes (Colautti & Lau 2015). Once parameterized with reciprocal transplant experiments, the VCG could be applied to reconstruct cline evolution in other invasive species and more generally in species experiencing novel and changing environments.

We have focused our analysis on flowering time of *Lythrum salicaria* in North America because we had observations from common garden experiments necessary to model effects of growing environment and to validate predicted phenological development times ($\psi$). However, the VCG analysis should be feasible for any trait that can be observed in (or inferred from) preserved specimens or other historical records provided (i) historical data are available to reconstruct environmental conditions at the time of collection, and (ii) a developmental model can be parameterized to account for different growing environments. In our case, population-by-environment interactions were not strong for the trait of interest, allowing our model to accurately recreate clines without knowledge of local genotype (Fig. 3). However, traits with stronger genotype- or population-by-environment interactions could be analyzed following the VCG if a more detailed model were available to correct for major demes or genotypes, which might be inferred using molecular genetic markers.

Convergent cline formation and stasis 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 over decades to centuries. The VCG analysis we introduce 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 window back through time but are vulnerable to loss without financial and academic support. Global efforts to manage biodiversity in the Anthropocene could benefit from a more predictive science of adaptive evolution in natural systems, but this will be difficult without continued support for natural history collections at a global scale.

**Data & Code availability**: A reproducible analysis, including all image sources and measurement data are available from the Dryad archive (DOI: Reviewer Note: zip files for Dryad are uploaded with manuscript and also available in public Repo on GitHub https://github.com/ColauttiLab/WuColauttiHerbarium).

**Acknowledgments:** We thank SCH Barrett, LH Rieseberg, CA Rushworth, KG Turner for manuscript feedback, and the following people for generously digitizing herbarium records for our study: M Pace & B Theirs (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), and A Liston (Oregon State U). This research was supported by NSERC Discovery grant to RIC.

**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

**References:**

Bell, G. & Gonzalez, A. (2009). Evolutionary rescue can prevent extinction following environmental change. *Ecology Letters*, 12, 942–948.

Blows, M.W. & Hoffmann, A.A. (2005). A Reassessment of Genetic Limits to Evolutionary Change. *Ecology*, 86, 1371–1384.

Chun, Y.J. (2011). Phenotypic plasticity of introduced versus native purple loosestrife: univariate and multivariate reaction norm approaches. *Biol Invasions*, 13, 819–829.

Colautti, R.I., Alexander, J.M., Dlugosch, K.M., Keller, S.R. & Sultan, S.E. (2017). Invasions and extinctions through the looking glass of evolutionary ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 20160031.

Colautti, R.I. & Barrett, S.C.H. (2010). Natural Selection and Genetic Constraints on Flowering Phenology in an Invasive Plant. *Int. J. Plant Sci.*, 171, 960–971.

Colautti, R.I. & Barrett, S.C.H. (2011). Population divergence along lines of genetic variance and covariance in the invasive plant *Lythrum salicaria* in eastern North America. *Evolution*, 65, 2514–2529.

Colautti, R.I. & Barrett, S.C.H. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342, 364–366.

Colautti, R.I., Eckert, C.G. & Barrett, S.C.H. (2010). Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proceedings of the Royal Society B: Biological Sciences*, 277, 1799–1806.

Colautti, R.I. & Lau, J.A. (2015). Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology*, 24, 1999–2017.

Colautti, R.I., Maron, J.L. & Barrett, S.C.H. (2009). Common garden comparisons of native and introduced plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications*, 2, 187–199.

Davis, C.C., Willis, C.G., Connolly, B., Kelly, C. & Ellison, A.M. (2015). 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.

Desmet, P. & Brouillet, L. (2013). 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.

Etterson, J.R., Franks, S.J., Mazer, S.J., Shaw, R.G., Gorden, N.L.S., Schneider, H.E., *et al.* (2016). Project Baseline: An unprecedented resource to study plant evolution across space and time. *American Journal of Botany*, 103, 164–173.

Ferrarese, E., Garono, R.J. & Schooler, S. (2009). *Assessment of Purple Loosestrife Biocontrol Agent Populations on the Columbia River*. Earth Design Consultants, Inc.

Franks, S.J., Hamann, E. & Weis, A.E. (2017). Using the resurrection approach to understand contemporary evolution in changing environments. *Evol Appl*, 11, 17–28.

Franks, S.J., Sim, S. & Weis, A.E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *PNAS*, 104, 1278–1282.

GBIF.org. (2017). GBIF: The Global Biodiversity Information Facility.

Gould, S.J. & Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature*, 366, 223–227.

Gräler, B., Pebesma, E. & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, 8, 204–218.

Grant, P.G. & Grant, R.G. (2014). *40 Years of Evolution: Darwin's Finches on Daphne Major Island*. Princeton University Press, Princeton, NJ.

Hendry, A.P. (2020). *Eco-evolutionary Dynamics*. Princeton University Press.

Hendry, A.P. & Kinnison, M.T. (1999). Perspective: The pace of modern life: measuring rates of contemporary microevolution. *Evolution*, 53, 1637–1653.

Hereford, J. (2009). A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The American Naturalist*, 173, 579–588.

Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol Ecol*, 26, 1101–1118.

Knapp, A.K., Smith, M.D., Hobbie, S.E., Collins, S.L., Fahey, T.J., Hansen, G.J.A., *et al.* (2012). Past, Present, and Future Roles of Long-Term Experiments in the LTER Network. *BioScience*, 62, 377–389.

Lang, P.L.M., Willems, F.M., Scheepens, J.F., Burbano, H.A. & Bossdorf, O. (2019). Using herbaria to study global environmental change, 221, 110–122.

Langlet, O. (1971). Two Hundred Years Genecology. *Taxon*, 20, 653–721.

Lindgren, C.J. & Walker, D. (2012). Growth Rate, Seed Production, and Assessing the Spatial Risk of Lythrum salicaria using Growing Degree-Days. *Wetlands*, 32, 885–893.

Lohmueller, K.E. (2014). The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet*, 10.

563 Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E. & Houston, T.G. (2012). An Overview of the
564 Global Historical Climatology Network-Daily Database. *Journal of Atmospheric &*
565 *Oceanic Technology*, 29, 897–910.

566 Merilä, J. & Hendry, A.P. (2014). Climate change, adaptation, and phenotypic plasticity: the
567 problem and the evidence. *Evolutionary Applications*, 7, 1–14.

568 Montague, J.L., Barrett, S.C.H. & Eckert, C.G. (2008). Re-establishment of clinal variation in
569 flowering time among introduced populations of purple loosestrife (Lythrum salicaria,
570 Lythraceae). *J. Evol. Biol.*, 21, 234–245.

571 Novembre, J. & Barton, N.H. (2018). Tread lightly interpreting polygenic tests of selection.
572 *Genetics*, 208, 1351–1355.

573 Oduor, A.M.O., Leimu, R. & Kleunen, M. van. (2016). Invasive plant species are locally adapted
574 just as frequently and at least as strongly as native plant species. *Journal of Ecology*, 104,
575 957–968.

576 Olden, J.D., LeRoy Poff, N., Douglas, M.R., Douglas, M.E. & Fausch, K.D. (2004). Ecological
577 and evolutionary consequences of biotic homogenization. *Trends in Ecology &*
578 *Evolution*, 19, 18–24.

579 Peischl, S., Dupanloup, I., Kirkpatrick, M. & Excoffier, L. (2013). On the accumulation of
580 deleterious mutations during range expansions. *Molecular Ecology*, 22, 5972–5982.

581 Perkins, T.A., Phillips, B.L., Baskett, M.L. & Hastings, A. (2013). Evolution of dispersal and life
582 history interact to drive accelerating spread of an invasive species. *Ecol. Lett.*, 16, 1079–
583 1087.

584 Savolainen, O., Pyhäjärvi, T. & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees.
585 *Annual Review of Ecology, Evolution, and Systematics*, 38, 595–619.

586 Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of
587 image analysis. *Nature Methods*, 9, 671–675.

588 Shamsi, S.R.A. & Whitehead, F.H. (1974). Comparative eco-physiology of *Epilobium hirsutum*
589 L. and *Lythrum salicaria* L.: II. Growth and development in relation to light. *Journal of*
590 *Ecology*, 62, 631–645.

591 Szűcs, M., Vahsen, M.L., Melbourne, B.A., Hoover, C., Weiss-Lehman, C. & Hufbauer, R.A.
592 (2017). Rapid adaptive evolution in novel environments acts as an architect of population
593 range expansion. *PNAS*, 114, 13501–13506.

594 Tjørve, K.M.C. & Tjørve, E. (2017). A proposed family of Unified models for sigmoidal growth.
595 *Ecological Modelling*, 359, 117–127.

596 US Geological Survey. (2019). *Species occurrence data for the Nation—USGS Biodiversity*
597 *Information Serving Our Nation (BISON)* (USGS Numbered Series No. 160). *Species*
598 *occurrence data for the Nation—USGS Biodiversity Information Serving Our Nation*
599 *(BISON)*, General Information Product. U.S. Geological Survey, Reston, VA.

600 Weider, L.J., Jeyasingh, P.D. & Frisch, D. (2018). Evolutionary aspects of resurrection ecology:
601 Progress, scope, and applications—An overview. *Evolutionary Applications*, 11, 3–10.

602 Williams, J.L., Kendall, B.E. & Levine, J.M. (2016). Rapid evolution accelerates plant
603 population spread in fragmented experimental landscapes. *Science*, 353, 482–485.

604 Willis, C.G., Ellwood, E.R., Primack, R.B., Davis, C.C., Pearson, K.D., Gallinat, A.S., *et al.*
605 (2017). Old Plants, new tricks: Phenological research using herbarium specimens. *Trends*
606 *in Ecology & Evolution*, 32, 531–546.

607

608

609