

**Title** Genome size variation and evolution during invasive range expansion in an introduced plant

**Running Title:** Genome size variation during invasion

**Authors:** F. Alice Cang<sup>1\*</sup>, Shana R. Welles<sup>1,2</sup>, Jenny Wong<sup>1</sup>, Maia Ziaeem<sup>1,3</sup>, Katrina M. Dlugosch<sup>1</sup>

**Correspondence:** F. Alice Cang, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Email: [alicecang@gmail.com](mailto:alicecang@gmail.com)

**Author Affiliations:**

<sup>1</sup>University of Arizona, Tucson, AZ, USA

<sup>2</sup>Utah Valley University, Orem, UT, USA

<sup>3</sup>Mills College, Oakland, CA, USA

**Abstract**

Plants demonstrate exceptional variation in genome size across species, and their genome sizes can also vary dramatically across individuals and populations within species. This aspect of genetic variation can have consequences for traits and fitness, but few studies attributed genome size differentiation to ecological and evolutionary processes. Biological invasions present particularly useful natural laboratories to infer selective agents that might drive genome size shifts across environments and population histories. Here, we test hypotheses for the evolutionary causes of genome size variation across 14 invading populations of yellow starthistle, *Centaurea solstitialis*, in California, USA. We use a survey of genome sizes and trait variation to ask: (1) Is variation in genome size associated with developmental trait variation? (2) Are genome sizes smaller toward the leading edge of the expansion, consistent with selection for ‘colonizer’ traits? Or alternatively, does genome size increase toward the leading edge of the expansion, consistent with predicted consequences of founder effects and drift? (3) Finally, are genome sizes smaller at higher elevations, consistent with selection for shorter development times? We found that 2C DNA content varied 1.21-fold among all samples, and was associated with flowering time variation, such that plants with larger genomes reproduced later, with lower

lifetime capitula production. Genome sizes increased toward the leading edge of the invasion, but tended to decrease at higher elevations, consistent with genetic drift during range expansion but potentially strong selection for smaller genomes and faster development time at higher elevations. These results demonstrate how genome size variation can contribute to traits directly tied to reproductive success, and how selection and drift can shape that variation. We highlight the influence of genome size on dynamics underlying a rapid range expansion in a highly problematic invasive plant.

**Keywords:** intraspecific genome size variation, invasive species, flow cytometry, yellow starthistle

### **Acknowledgements**

We thank Abreeza Zegeer for advice on greenhouse experimental set up and management. We thank Paula Campbell and John Finch from the University of Arizona Cytometry Core Facility, and Anthony Baniaga for their assistance preparing flow cytometry protocols for genome size estimation. We would also like to thank Joseph Braasch for supplying dates of population establishment and helpful feedback. We are most grateful to Rhiannon Bauer, Alexus Cazares, Shelby Dolgaard and Alex Skomro for their indispensable help with collecting common garden data. Funding was provided by USDA grant #2015-67013-23000 and NSF grant #1750280 to KMD.

**Data for this study are available at:** *to be completed after manuscript is accepted for publication.*

### **Ethics**

This manuscript has not been submitted elsewhere and is not under consideration for publication in another journal or book.

### **Conflict of Interest Statement**

The authors declare no conflict of interest.

## Abstract

Plants demonstrate exceptional variation in genome size across species, and their genome sizes can also vary dramatically across individuals and populations within species. This aspect of genetic variation can have consequences for traits and fitness, but few studies have attributed genome size differentiation to ecological and evolutionary processes. Biological invasions present particularly useful natural laboratories to infer selective agents that might drive genome size shifts across environments and population histories. Here, we test hypotheses for the evolutionary causes of genome size variation across 14 invading populations of yellow starthistle, *Centaurea solstitialis*, in California, USA. We use a survey of genome sizes and trait variation to ask: (1) Is variation in genome size associated with developmental trait variation? (2) Are genome sizes smaller toward the leading edge of the expansion, consistent with selection for ‘colonizer’ traits? Or alternatively, does genome size increase toward the leading edge of the expansion, consistent with predicted consequences of founder effects and drift? (3) Finally, are genome sizes smaller at higher elevations, consistent with selection for shorter development times? We found that 2C DNA content varied 1.21-fold among all samples, and was associated with flowering time variation, such that plants with larger genomes reproduced later, with lower lifetime capitula production. Genome sizes increased toward the leading edge of the invasion, but tended to decrease at higher elevations, consistent with genetic drift during range expansion but potentially strong selection for smaller genomes and faster development time at higher elevations. These results demonstrate how genome size variation can contribute to traits directly tied to reproductive success, and how selection and drift can shape that variation. We highlight the influence of genome size on dynamics underlying a rapid range expansion in a highly problematic invasive plant.

## Introduction

Plants demonstrate some of the greatest variation in genome size among eukaryotes (Bennetzen *et al.* 2005) and the potential ecological and evolutionary consequences of this widespread variation remain an open question. At the most basic level, we expect that large genomes require greater cell volumes and longer replication times, and comparative analyses across plants bear out the generality of this relationship, along with associated reductions in stomatal density and increased seed mass (Beaulieu *et al.* 2007a; Beaulieu *et al.* 2008). These differences suggest potential costs associated with maintaining large genomes, as greater cellular volumes and longer cell cycles might constrain functional traits related to growth, reproduction and dispersal, including longer minimum generation times (Grotkopp *et al.* 2004), slower rates of photosynthesis (Knight *et al.* 2005) and smaller plant size (Carta and Peruzzi 2016), among others (e.g. Kenton *et al.* 1986; Francis *et al.* 2008; Vesely *et al.* 2020). While there has been

some success in identifying these effects at the cellular level and subsequent trait changes (Knight and Beaulieu 2008; Hodgson *et al.* 2010), population consequences and fitness effects in particular environments remain less understood (Gaut and Ross-Ibarra 2008; Whitney *et al.* 2010; Mei *et al.* 2017; Bilinski *et al.* 2018).

Contrary to traditional understanding of genome size as a stable species-level trait, genome structure can vary dramatically across individuals and populations and over time, even on very short time scales. The predominant molecular mechanisms of genome expansion are whole genome or whole chromosome duplications, and transposable element (TE) proliferation within a ploidy level (Feschotte and Pritham 2007; Tenaillon *et al.* 2010; Lisch 2013). Conversely, unequal homologous recombination or illegitimate recombination drive losses of DNA content (Bennetzen *et al.* 2005). Some of the clearest evidence for the mutability of genome content and the processes that influence genome size has been observed in crop systems and wild crop relatives. Some of the earliest evidence emerged from studies of wild barley, in which variation in insertion patterns of a single family, BARE-1, has contributed up to ~5% of genome size variation (Kalendar *et al.* 2000). In maize, the most common TE families that make up nearly ~70% of the genome have average insertion dates of less than 1 Mya, with some inserting as recently as 560 kya (Baucom *et al.* 2009). These recent bursts of TE proliferation are responsible for wide-ranging genome size variation in maize, and repeat content explains a >20% difference in genome size between the *Palomero* landrace and the B73 inbred line (Vielle-Calzada *et al.* 2009). Likewise, differences in TE purging are strongly correlated with genome size variation in maize, where significant loss of TE content was observed after only a few generations in experimental maize lines (Roessler *et al.* 2019). Similar findings in flax have identified a comparable scale of variation, as well as inducible changes to genome size within one generation (Cullis 2005), suggesting relatively few meiotic events are required to generate dramatic variation among closely related individuals. In cultivated rice, recent transposition of LTR-RT transposable element families are responsible for genome size doubling, on the same scale of variation as whole genome duplications (Piegú *et al.* 2006). Work in select model systems has shown similar findings, including in *Arabidopsis* (Schmuths *et al.* 2004; Davison *et al.* 2007) and *Drosophila* species (Petrov 2003; Ellis *et al.* 2014).

Few studies have been able to clearly attribute genome size variation to ecological and evolutionary processes. The *Arabidopsis* (Lockton *et al.* 2008; Lockton and Gaut 2010) and *Drosophila* genetic model systems (Gonzalez *et al.* 2008a; Gonzalez *et al.* 2008b) have leveraged abundant sequence data and have identified signatures of both selection and drift in families of TEs known to contribute to variation in genome size. Again, crop systems also provide evolutionary insights, as in the genus *Oryza*, where TE variation was identified as the driver of divergence between different species, indicating that mechanisms generating genome size variation can also have important macroevolutionary implications (Zhang and Gao 2017). In wild barley populations, BARE-1 retroelement content varied across elevational and aridity gradients, suggesting an association with environmental conditions (Kalendar *et al.* 2000). Multiple systems have identified potential selective effects of elevation on genome size, potentially due to differences in growing season and the fitness effects of development time, as in maize (Bilinski *et al.* 2018), *Corchorus olitorius* (Benor *et al.* 2011), *Lagenaria siceraria* (Achigan-Dako *et al.* 2008), as well as a small number of non-crop systems such as *Dactylis glomerata* (Creber *et al.* 1994), and *Arachis duranensis* (Temsch and Greilhuber 2001), though other results are conflicting (Lysak *et al.* 2000; Oney-birol and Tabur 2018; Savas *et al.* 2019). Taken together, these studies suggest that intraspecific genome size variation arises because the molecular mechanisms that influence genome size and subsequent structural variation are subject to the same evolutionary processes as other mutations. Subsequent accumulation of population-level differences can be ecologically relevant and function as important sources of variation upon which selection can act. Yet, there is still a dearth of knowledge about the fitness effects of genome size variation in natural populations, and in non-model systems.

Biological invasions present particularly useful natural laboratories to examine ecological consequences of intraspecific genome size variation and to infer selective agents that might drive genome size shifts. If genome size imposes developmental constraints, a potential invader may benefit from reduced genome sizes that promote fast generation time and high reproductive rates, which are traits associated with a colonizer life history (Baker 1974). Thus, genome size-associated traits that equip individuals to disperse, survive and reproduce in marginal habitats, may select for smaller genomes. In large multi-species comparisons, reported incidence of naturalization was more likely among species with smaller genomes, suggesting

potential advantages during early stages of invasion (Pysek *et al.* 2023). This may account for the overrepresentation of species with small genomes among weedy taxa in broad-scale surveys of plant invaders globally (Kubesova *et al.* 2010; Pandit *et al.* 2014) and across the US (Kuester *et al.* 2014). A comparative phylogenetic study of pines identified relationships between small genomes and “weedy” traits, such as small seed size, short minimum generation time and fast relative growth rate, while also linking genome size to invasiveness as a character (Grotkopp *et al.* 2004). Importantly, invaded systems allow inter- and intra-population comparisons of introductions with their source populations, to infer contemporary and rapid evolution of genome size. Direct comparisons between introduced populations and known sources are relatively few in number (Crosby *et al.* 2014; Pysek *et al.* 2018) and methodological issues with genome size estimation can produce contested results (Lavergne *et al.* 2009; Martinez *et al.* 2018). Identifying evidence for selection on genome size during colonization requires systems in which the history of expansion is well-documented, ecologically relevant traits are known, and robust genome size estimation methods are used.

In contrast to natural selection, neutral population genetic processes might also influence genome size variation. If processes such as TE proliferation generate genetic variants that are neutral or only weakly deleterious, genetic drift may allow sequences that increase average genome size to persist and spread. This is especially important in invading populations, where range expansion can diminish the effective population size of successive founding events at the leading edge (Slatkin and Excoffier 2012; Braasch *et al.* 2019), and mutations arising in advancing populations can ‘surf’ to higher frequency and broader geographic scales (Klopfenstein *et al.* 2005; Gralka *et al.* 2016), even if these variants are disadvantageous (Peischl *et al.* 2013). Thus, if drift and founder effects are stronger than selection, we expect genome sizes to increase as an invasion expands across space.

Here, we investigate evolutionary causes of genome size variation in the invasive species yellow starthistle, *Centaurea solstitialis* L., in the Asteraceae (hereafter, YST). YST is an annual, obligately outcrossing, diploid thistle ( $2n=16$ , Heiser and Whitaker 1948; Widmer *et al.* 2007; Öztürk *et al.* 2009), and previous work has suggested that genome size is variable within its invaded ranges (Irimia *et al.* 2017). YST is invasive in the United States, where it is particularly

problematic in California and invading populations are now distributed across more than 12 million acres of land (Maddox and Mayfield 1985; Pitcairn *et al.* 2006). YST likely originated in the eastern Mediterranean before an ancient expansion into Eurasia and western Europe (Barker *et al.* 2017). Historical records indicate that it was introduced to South America from Spain and to the US from Chile via imports of alfalfa, with the earliest recorded occurrence in the San Francisco Bay area in 1869 (Pitcairn *et al.* 2006). This recent colonization history has been confirmed by genetic studies (Dlugosch *et al.* 2013; Eriksen *et al.* 2014; Barker *et al.* 2017). We know from this previous work that Californian populations are derived from source populations in western Europe (Dlugosch *et al.* 2013; Barker *et al.* 2017) and have evolved shifts in growth rate and flowering time (Dlugosch *et al.* 2015), which are traits associated with genome size variation in other systems (eg. Grotkopp *et al.* 2004; Carta and Peruzzi 2015; Bilinski *et al.* 2018). We also know that the invasion has crossed large elevational and environmental gradients (Gerlach 1997; Pitcairn *et al.* 2006; Braasch *et al.* 2019), which provide an opportunity to examine the influence of environmental clines on genome size variation, decoupled from invasion history. In particular, evidence from other plant systems (eg. Achigan-Dako *et al.* 2008; Benor *et al.* 2011; Bilinski *et al.* 2018) suggests higher elevations in particular may impose selection for smaller genome sizes.

We leverage understanding of the expansion history of YST in California to test predictions about how genome size variation might have evolved under different evolutionary processes. If large genome size constrains growth and reproduction as observed in other systems, we expect selection to favor smaller genome sizes during the process of range expansion, and at high elevations regardless of expansion history. Alternatively, if neutral processes dominate, we expect genome sizes to increase during range expansion. We combine a survey of genome sizes estimated by flow cytometry with trait data collected from a common garden experiment to ask: (1) Is variation in genome size associated with developmental trait variation? (2) Are genome sizes smaller toward the leading edge of the expansion, consistent with selection for ‘colonizer’ traits? Or alternatively, does genome size increase toward the leading edge of the expansion, consistent with expectations of founder effects and drift? (3) Finally, are genome sizes associated with elevational clines that are decoupled from the direction of expansion?

## Materials and Methods

### *Plant material*

YST is distributed patchily across California, and local populations within this invasion have been observed to vary in terms of both traits (Dlugosch *et al.* 2015) and effective population size Braasch *et al.* (2019). Seeds were collected from 14 sites across California (Fig. 1, Table 1) in August 2016 along a linear transect at each site, with maternal plants >1m apart. Collections include populations from the coastal San Francisco Bay area which is the putative site of introduction to California, the Central Valley where the invasion is most severe and the Sierra Nevada mountain range, which includes the leading edge of the Californian expansion.

### *Common garden*

We collected common garden data from 13-30 individuals from each of the 14 populations (N=312; Table 1). We germinated seeds on the surface of moist potting soil (3:2:1 ratio of Sunshine Mix #3 soil, vermiculite and 20 grit silica sand) under fluorescent lights and 12 hour days in December 2016 and recorded germination date daily. We germinated multiple seeds from the same maternal plant and kept the first germinating individual. We transplanted all growing plants into 410ml Deepots (Steuwe and Sons, Inc, USA) in January 2017 when they were 5 weeks old and grew them in a greenhouse at the University of Arizona in Tucson, AZ, USA. Once in the greenhouse, we randomly assigned one individual from each population to each of 30 different blocks. Additional individuals from two different maternal plants from ARN and SAN were transplanted into three and four blocks, respectively (Table 1). Plants were watered daily using an automatic drip watering system and maintained through senescence. At the rosette stage, prior to reproduction, we measured leaf number, length of longest leaf and width of longest leaf to estimate a size index and early growth rate (as in Dlugosch *et al.* 2015). The size index was calculated as (leaf number\*(maximum leaf length\*maximum leaf width) $^{1/2}$ ), which correlates linearly with biomass in this species (Dlugosch *et al.* 2015). We calculated linear growth rates using our size index and number of days since germination. The common garden was checked daily to record days to the initiation of bolting and days until the first flower. Plants were harvested when most individuals had died, after which we counted the total number

of flowering heads. We dried harvested plants overnight at 60°C and weighed aboveground dry biomass.

#### *Genome size estimation*

We estimated genome size by flow cytometry using the FACSCanto II instrument (BD Biosciences, San Jose, CA, USA) equipped with a blue (488-nm), air-cooled, 20-mW solid state laser and a red (633-nm) 17-mW helium neon laser for UV excitation. Sample preparation followed a modified two-step protocol by Doležel *et al.* (2007) as follows. We chose *Raphanus sativus* (2C=1.11pg) provided by the Institute of Experimental Botany (Prague, Czech Republic) as our internal standard, as it has a close but non-overlapping genome size with the previously reported nuclear genome size of YST (2C=1.74pg; Bancheva and Greilhuber 2006). Nuclear suspensions were prepared by chopping 50-60 mg each of YST and *R. sativus* fresh tissue with 0.5 ml of ice-cold Otto I buffer in a Petri dish on top of ice, using a sharp razor blade. The suspension was filtered first through gauze and then again through 18 µm Nylon mesh. Nuclear suspensions in Otto I buffer are considered relatively stable (Doležel *et al.* 2007) and samples were left covered and on ice until immediately before analysis by flow cytometry, at which point we added 0.77ml of Otto II buffer, 29 µL of 1 mg/mL propidium iodide (PI) stain and 7.5 µL of 1 mg/mL RNase. The sample was then gently vortexed and left for at least 5 minutes on ice in a covered container. YST genome sizes were calculated according to Doležel and Bartos (2005). We recorded 3 estimates of genome size on different days for a subset of individuals (N=24), including 1-3 individuals from 13 of our 14 populations. We obtained a single estimate of genome size for 13-28 additional individuals from across those 14 populations, for a total 333 individuals for which we have at least 1 genome size estimate (Table 1).

#### *Genome size correction*

Preliminary analyses of the genome size data indicated that the date of flow cytometry measurement had a significant effect on estimated genome size (see Results, Table 2). This trend might reflect variation induced by developmental changes in the plant as it ages or changes in the flow cytometer over time (Doležel and Bartoš 2005; Doležel *et al.* 2007). To account for potential variation due to measurement error associated with the estimation date, we used the

best fitting linear model predicting genome size (Table 2) to inform date-corrected estimates that we included in subsequent trait analyses.

#### *Statistical analyses*

The growth and flowering traits that we collected are likely to be correlated as part of plant development, and so we used a principal component analysis (PCA) to identify major axes of variation in these traits using base R (R Core Team, 2021). Coordinates of PC1 and PC2 were then used as response variables in linear models of the effect of genome size on traits. We fit separate linear models to predict PC1 and PC2 from fixed effects of corrected genome size, greenhouse position, days until harvesting, and interactions among all fixed effects. Greenhouse position was treated as a fixed ordinal variable, reflecting the distance from the evaporative cooling system at one end of the greenhouse, due to a temperature gradient of  $\sim 8^{\circ}\text{C}$  across our experiment. For both analyses, model selection was performed using analyses of variance (ANOVA) to identify if the removal of non-significant effects ( $p > 0.1$ ) significantly improved the model.

We fit linear models to test our hypotheses that genome size declines toward the leading edge of invasion and at high elevations within the Californian invasion. Corrected genome size (N=333) was modeled as a function of fixed effects of age of source population, elevation, and position in the greenhouse. We used the Jepson Online Herbarium (<http://ucjeps.berkeley.edu/>) for *C. solstitialis* in California to determine population age as in Braasch *et al.* (2019). We used the earliest occurrence record in the county of each source population, or the record for a neighboring county if the collection site was closer to an older occurrence record, as its colonization date. We found elevation using Google Earth records from latitude and longitude of collection locations. The analysis was repeated for genome size estimates that had 3 replicates (N=24), using the mean value across replicates.

## **Results**

Our survey of populations in the Californian invasion identified intraspecific 2C genome size variation of 1.21 fold across individuals, ranging from 1.58-1.85pg (1.47 - 1.77 GB), with a

mean of 1.72pg (1.68 GB) across all genome size measurements (Fig. 2). In our reduced dataset of only individuals with repeated estimates, we determined the means across each individual's estimates and recovered the same range of genome size from 1.58-1.85pg (Suppl. Fig. S1), with a slightly greater mean of 1.74pg (1.70 Gbp). We also found a high degree of within-population variation in genome size, such that most populations included genome sizes spanning the range of the dataset. For example, the coastal Bay area populations MAR16 and NAP had the greatest intrapopulation range of 1.58-1.78pg (1.55-1.74 GB) and 1.60-1.80pg (1.56-1.76 GB), respectively, and the Central Valley population RB had the narrowest range in genome sizes at 1.66 - 1.76pg (1.62-1.72 GB).

The best-fitting linear model explaining variation in genome size identified significant negative effects of population age, such that more recently founded populations had higher genome sizes, consistent with a scenario of reduced efficacy of selection in recently established populations (Fig 3a, Table 2). The effect of elevation had a marginally significant, negative relationship with genome size, consistent with selection for faster growth and early reproduction at higher elevations, where growing seasons tend to be shorter (Fig 3c, Table 2). We repeated the analysis of genome size with 24 individuals for which we had 3 replicate estimates. We again found a significant negative relationship with population age (Fig 3b); however, there was no significant relationship with elevation in the reduced dataset, and no effect of genome size estimation date. Across the full dataset of genome sizes, we identified a weakly negative but significant effect of estimation date, wherein flow cytometry measurements later in the growing season were significantly smaller ( $\beta=-0.00031, p<0.001$ ). For models assessing trait variation, we applied the following correction to our raw estimates:

*Corrected genome size =*

$$\text{Estimated genome size } -(-3.061 \times 10^{-4})(\text{Days to genome size estimation})$$

Among growth and flowering traits, the first two principal components explained 67.3% of total variance, with PC1 and PC2 accounting for 40.9% and 26.4% variance, respectively (Fig. 4). Principal component analysis indicated that PC1 was strongly associated with growth rates and aboveground dry biomass, while PC2 was associated with bolting time, flowering time, and flower number at harvest (Fig. 4). Higher PC1 values corresponded to faster growth and larger

overall size. Higher PC2 values corresponded to longer time to bolting and flowering, and fewer flowers produced overall. PC1 and PC2 coordinate values were used as measures of growth and development metrics for subsequent analyses of trait variation.

In total, we collected both flow cytometry and trait data from 312 individuals. A linear model explaining variation in PC1 scores included non-significant main effects of corrected genome size, with significant interactions between genome size and greenhouse position (Table 3). A partial residual plot of PC1 versus genome size shows a positive slope (Fig 5a), despite a negative coefficient predicted by the linear model on the effect of genome size alone (Table 3). This discrepancy occurred because the effect of genome size on growth-related traits was dependent on where in the greenhouse an individual was grown (Fig. 6). There were no significant effects of genome size on PC1 in positions 1-6, where environmental conditions were warmer, but positions 7 and 8 demonstrated significant associations with genome size where conditions were cooler (Suppl. Table 1). We also found the direction of the relationship between genome size and growth-related traits was more positive in blocks that were cooler, as in the case of position 9 (Fig 6). A linear model explaining variation in PC2 indicated a significant positive effect of corrected genome size ( $\beta=6.476, p=0.001$ ) and negative effects of greenhouse position and its interaction with harvesting time (Table 4, Suppl. Table 2). Genome size was significantly, positively correlated with the timing of reproductive events, wherein higher PC2 axis values indicate a longer time to bolting and flowering, as well as lower capitula production (Fig. 5b), consistent with the prediction that genome replication costs and longer cell cycles may delay reproduction.

## Discussion

Intraspecific homoploid genome size variation has been demonstrated in several plant systems, but the ecological and evolutionary factors shaping this variation, and its potential contribution of this genetic variation to species invasions, remain unclear. We conducted a survey of genome size and trait variation in YST populations across its invasion in California to test alternative eco-evolutionary hypotheses explaining divergence in genome size during range expansion. We found that the 2C DNA content varied 1.21-fold among all California samples, suggesting widespread and extensive genome size variation across this invasion. Genome size

variation was associated with flowering time, such that plants with larger genomes reproduced later, and with lower lifetime capitula production. Genome sizes increased toward the leading edge of the invasion, but tended to decrease at higher elevations, consistent with genetic drift during range expansion but potentially strong selection for smaller genomes and faster development time at higher elevations.

Surveys of intraspecific genome size variation in other systems have documented both broad-scale variation and conserved genome sizes across populations. For example, natural populations included 1.7-fold variation in diploid genome size of *Brachypodium distachyon* (Oney-Birol and Tabur 2018), and over 2-fold variation in *Synura petersenii* (Certnerova and Skaloud 2020). Conversely, genome size can be a very stable species-level trait with little observed variation, despite widespread population sampling (eg. Lysak *et al.* 2000; Savas *et al.* 2019). With a 1.21-fold 2C range, YST demonstrated moderate variation on par with a previous survey in the system (Irimia *et al.* 2017) and with observations in several other systems including *Dactylis glomerata* (Creber *et al.* 1994), *Arachis duranensis* (Temsch and Greilhuber 2001), *Lagenaria siceraria* (Achigan-Dako *et al.* 2008) and *Phragmites australis* (Pysek *et al.* 2018). The variation we observed in YST is also similar to systems in which significant associations between genome size and plant traits (Pysek *et al.* 2018), or environmental variables (Benor *et al.* 2007; Achigan-Dako *et al.* 2008; Bilinski *et al.* 2018) have been demonstrated. Our means of triplicate genome size estimates also span this range of variation, and bracket previous single estimates (Bancheva and Greilhuber 2006; Miskella 2014; Carev *et al.* 2017), indicating that measurement errors are unlikely to explain the 1.21-fold variation observed across our samples (Greilhuber 2005).

Notably, we detected an influence of measurement date on genome size estimates, in which later estimates tended to yield smaller genome sizes. Flow cytometry measurements that underestimate nuclear size can occur when DNA staining is inhibited (Doležel *et al.* 2007). Genome size measurements were taken over five months, during which the plants progressed from rosette formation in early development, through early senescence. Phenolic concentrations in the cytosol can vary among tissue types and across developmental stages (Witzell *et al.* 2003; Wam *et al.* 2017), some of which are known to inhibit the staining of DNA (Doležel *et al.* 2007).

Factors that could interfere with these measurements differ across species (Doležel *et al.* 2007) and methodology (Price *et al.* 2000; Noirot *et al.* 2000; Noirot *et al.* 2003; Noirot *et al.* 2005; Loureiro *et al.* 2007), which may compromise large-scale comparisons that take many measurements over extended periods of time. Although previous work in YST found no overall differences in genome size among different parts of its global distribution (Irimia *et al.* 2017), an analysis accounting for changes in measurement over time does find evidence of significant differences between native and invaded ranges (Cang and Dlugosch 2022, *bioarxiv*). Fortunately, our relatively large sample size allowed us to identify this change over time and most importantly, use our linear model explaining genome size variation to quantify the effect and remove it from subsequent analyses of trait variation.

Our finding that larger genomes were associated with later flowering and lower flower production joins a growing body of evidence pointing to potential developmental costs of maintaining large genomes. Across a diversity of species, genome size has been correlated with a suite of characteristics, including plant size (Munzbergova 2009; Carta and Peruzzi 2016), growth rate (Fridley and Craddock 2015), pollen tube growth (Reese and William 2019), seed size (Beaulieu *et al.* 2007), flowering time (eg. Benor *et al.* 2011; Jian *et al.* 2017), and rates of photosynthesis (Roddy *et al.* 2020). Comparative phylogenetic approaches to identifying the effects of genome size on traits have found that genome size can be a strong predictor of seed mass (Knight and Beaulieu 2008), which may mediate the apparent effect of genome size on later developing traits (Grotkopp *et al.* 2004).

Strikingly, the magnitude and direction of genome size effects on growth-related traits in our study were dependent on local environmental conditions. Our linear model demonstrated a significant effect of the interaction between greenhouse position and genome size on growth, in which genome size had a significant positive association with growth in experimental blocks that were in cooler locations. The nature of this interaction was predominantly driven by individuals with small genomes that exhibited slower growth and produced less biomass overall under cooler growing conditions. A similar effect has also been identified in *Phragmites australis* (Meyerson *et al.* 2020). These interactions suggest that the effect of selection on genome size, and the subsequent relationship between genome size and the traits it influences, is likely to vary across

environments. For example, field surveys of an experimental grass community in the UK found that species with large genomes were underrepresented in nutrient-poor conditions with lower productivity, but successfully competed with small genome species and produced more biomass in N+P supplemented plots (Guignard *et al.* 2016). Principal component analysis of traits in our study identified that major axes associated with growth and reproduction varied independently from each other, but genome size influenced variation across both axes. These relationships suggest that effects of genome size on overall plant performance and fitness are complex, and can manifest across multiple independent traits and be environmentally-dependent. A critical consideration for additional experiments is that the apparent scale of genome size effects on trait variation may depend on the environments under which they are measured. Variation in the strength of these effects could then be difficult to quantify without explicit testing under varying environments, and associations identified under experimental conditions might differ from the effects of genome size variation in the field.

We predicted that selection would favor smaller genome sizes in habitats that favor faster development times. There is increasing evidence of correlations between genome size and a variety of environmental conditions, such as elevation (Achigan-Dako *et al.* 2008; Benor *et al.* 2011; Diez *et al.* 2013; Carta and Peruzzi 2015; Bilinski *et al.* 2018), temperature seasonality (Knight and Ackerly 2002; Paule *et al.* 2018; Qiu *et al.* 2019), atmospheric carbon (Franks *et al.* 2012), precipitation (Knight and Ackerly 2002; Carta and Peruzzi 2015), heavy metal pollution (Vidic *et al.* 2009; Temsch *et al.* 2010) and soil nitrogen (Kang *et al.* 2015; Guignard *et al.* 2016). We found marginally significant evidence for small genome sizes at high elevations in our full dataset. However, the marginally significant relationship with elevation disappeared in the reduced dataset, as did the effect of date of genome size estimation. These replicates were taken early in YST development for all individuals in the reduced dataset, and so variation in age-specific cytosolic compound concentrations likely exerted less influence on estimation precision. The relationship between genome size and elevation is thought to be mediated by selection for rapid onset of flowering due to short growing seasons at high elevations, as appears to be the case in maize (Bilinski *et al.* 2018). We do observe a significant negative correlation between flowering time and elevation, such that stronger direct selection on flowering time might explain weaker patterns of genome size differentiation, given that genome size differences

explain only a portion of flowering time trait variation. Our linear models cannot establish cause and effect between genome size and flowering time, however, and the fitness costs of genome size variation should be quantified in the field across elevational gradients.

In contrast, we did not find evidence that genome size declines during range expansion, as we predicted if selection favored fast development times in invaders. Rather, YST genome sizes increased toward the leading edge of the expansion, a pattern more consistent with founder effects and genetic drift contributing to the accumulation of larger genomes in populations at the leading edge. Repeated analysis of genome size with individuals for which we had three replicate estimates confirmed a significant negative relationship with population age. Previous work in this system demonstrated that more recently founded populations have lower effective population sizes, consistent with stronger genetic drift (Braasch *et al.* 2019). Recent theoretical work on the formation of range limits predicts that mutational load resulting from accumulation of deleterious variants during successive founding events can reduce population expansion across an environmental gradient, especially when mutations are not severely deleterious (Henry *et al.* 2015). A survey of genome sizes in European beech recovered similar findings, in which genome sizes increased towards range margins (Paule *et al.* 2018). This has particular importance for our understanding of the evolution of range limits in introduced species, as relatively strong genetic drift could allow mildly deleterious increases in genome size to accumulate and slow expansion at the leading edge. Geographic patterns of mutation accumulation in *Arabidopsis lyrata* support this prediction, in which signatures of mutational load increased towards range edges, with associated effects on individual plant performance and population growth (Willi *et al.* 2018). Therefore, smaller effective population sizes is one potential explanation for the persistence of larger genomes at the edge of expansion in spite of potential reproductive costs, and why a putatively adaptive relationship with elevation is weak. Disentangling the direction and magnitude of these potentially opposing influences on genome size will require direct estimates of selection on traits and underlying genetic polymorphisms, as well as expected population allele frequencies under neutral conditions of the sequences contributing to genome size variation.

Successful management of species invasions and prevention of future spread depends on understanding the factors that shape the dynamics of range expansion. There is a long history in invasion biology of seeking traits that are associated with invaders, but we have only recently begun to investigate structural genomic variation (Kuester *et al.* 2014; Schrader *et al.* 2014; Stapley *et al.* 2015; Gourbet *et al.* 2017; Bilinski *et al.* 2018; Pysek *et al.* 2018; Baduel *et al.* 2019). Our data contribute to increasing evidence that intraspecific genome size variants are common and could generate genetic and phenotypic variation on very short timescales, which may be important to establishment and range edge formation during invasions.

### **Acknowledgements**

We thank Abreeza Zegeer for advice on greenhouse experimental set up and management. We thank Paula Campbell and John Finch from the University of Arizona Cytometry Core Facility, and Anthony Baniaga for their assistance preparing flow cytometry protocols for genome size estimation. We would also like to thank Joseph Braasch for supplying dates of population establishment and helpful feedback. We are most grateful to Rhiannon Bauer, Alexus Cazares, Shelby Dolgaard and Alex Skomro for their indispensable help with collecting common garden data. Funding was provided by USDA grant #2015-67013-23000 and NSF grant #1750280 to KMD.

**Data for this study are available at:** *to be completed after manuscript is accepted for publication.*

### **Literature cited**

Achigan-Dako, E. G., Fuchs, J., Ahanchede, A., & Blattner, F. R. (2008). Flow cytometric analysis in *Lagenaria siceraria* (Cucurbitaceae) indicates correlation of genome size with usage types and growing elevation. *Plant Systematics and Evolution*, 276(1–2), 9–19. <https://doi.org/10.1007/s00606-008-0075-2>

Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., & Colot, V. (2019). Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nature Communications*, 10(1), 1–10. <https://doi.org/10.1038/s41467-019-13730-0>

Baker, H. G. (1974). The evolution of weeds. *Annual Review of Ecology and Systematics*, 5(1), 1–24.

Bancheva, S., & Greilhuber, J. (2006). Genome size in Bulgarian *Centaurea* s.l. (Asteraceae). *Plant Systematics and Evolution*, 257(1–2), 95–117. <https://doi.org/10.1007/s00606-005-0384-7>

Barker, B. S., Andonian, K., Swope, S. M., Luster, D. G., & Dlugosch, K. M. (2017). Population genomic analyses reveal a history of range expansion and trait evolution across the native and invaded range of yellow starthistle (*Centaurea solstitialis*). *Molecular Ecology*, 26(4), 1131–1147. <https://doi.org/10.1111/mec.13998>

Baucom, R. S., Estill, J. C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R. P., SanMiguel, P. J., & Bennetzen, J. L. (2009). Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genetics*, 5(11), e1000732. <https://doi.org/10.1371/journal.pgen.1000732>

Beaulieu, J. M., Leitch, I. J., & Knight, C. A. (2007). Genome size evolution in relation to leaf strategy and metabolic rates revisited. *Annals of Botany*, 99(3), 495–505. <https://doi.org/10.1093/aob/mcl271>

Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., & Knight, C. A. (2008). Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist*, 179(4), 975–986. <https://doi.org/10.1111/j.1469-8137.2008.02528.x>

Beaulieu, J. M., Moles, A. T., Leitch, I. J., Bennett, M. D., Dickie, J. B., & Knight, C. A. (2007). Correlated evolution of genome size and seed mass. *New Phytologist*, 173(2), 422–437. <https://doi.org/10.1111/j.1469-8137.2006.01919.x>

Bennetzen, J. L. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of Botany*, 95(1), 127–132. <https://doi.org/10.1093/aob/mci008>

Benor, S., Fuchs, J., & Blattner, F. R. (2011). Genome size variation in *Corchorus olitorius* (Malvaceae s.l.) and its correlation with elevation and phenotypic traits. *Genome*, 54(7), 575–585. <https://doi.org/10.1139/g11-021>

Bilinski, P., Albert, P. S., Berg, J. J., Birchler, J. A., Grote, M. N., Lorant, A., Quezada, J., Swarts, K., Yang, J., & Ross-Ibarra, J. (2018). Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLOS Genetics*, 14(5), e1007162. <https://doi.org/10.1371/journal.pgen.1007162>

Braasch, J., Barker, B. S., & Dlugosch, K. M. (2019). Expansion history and environmental suitability shape effective population size in a plant invasion. *Molecular Ecology*, 28(10), 2546–2558. <https://doi.org/10.1111/mec.15104>

Cang, A., & Dlugosch, K. M. (2022). Ecological effects of genome size in yellow starthistle (*Centaurea solstitialis*) vary between invaded and native ranges. *bioRxiv*, 2022-10. <https://doi.org/10.1101/2022.10.25.513778>

Carev, I., Ruščić, M., Skočibušić, M., Maravić, A., Siljak-Yakovlev, S., & Politeo, O. (2017). Phytochemical and cytogenetic characterization of *Centaurea solstitialis* L. (Asteraceae) from Croatia. *Chemistry & Biodiversity*, 14(2), e1600213. <https://doi.org/10.1002/cbdv.201600213>

Carta, A., & Peruzzi, L. (2016). Testing the large genome constraint hypothesis: plant traits, habitat and climate seasonality in Liliaceae. *New Phytologist*, 210(2), 709–716. <https://doi.org/10.1111/nph.13769>

Čertnerová, D., & Škaloud, P. (2020). Substantial intraspecific genome size variation in golden-brown algae and its phenotypic consequences. *Annals of Botany*, 126(6), 1077–1087. <https://doi.org/10.1093/aob/mcaa133>

Creber, H. M. C., Davies, M. S., Francis, D., & Walker, H. D. (1994). Variation in DNA C value in natural populations of *Dactylis glomerata* L. *New Phytologist*, 128(3), 555–561.

Crosby, K., Stokes, T. O., & Latta, R. G. (2014). Evolving California genotypes of *Avena barbata* are derived from multiple introductions but still maintain substantial population structure. *PeerJ*, 2, e633. <https://doi.org/10.7717/peerj.633>

Cullis, C. A. (2005). Mechanisms and control of rapid genomic changes in flax. *Annals of Botany*, 95(1), 201–206. <https://doi.org/10.1093/aob/mci013>

Davison, J., Tyagi, A., & Comai, L. (2007). Large-scale polymorphism of heterochromatic repeats in the DNA of *Arabidopsis thaliana*. *BMC Plant Biology*, 7(1), 1–15. <https://doi.org/10.1186/1471-2229-7-44>

Díez, C. M., Gaut, B. S., Meca, E., Scheinvar, E., Montes-Hernandez, S., Eguiarte, L. E., & Tenaillon, M. I. (2013). Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*, 199(1), 264–276. <https://doi.org/10.1111/nph.12247>

Dlugosch, K. M., Cang, F. A., Barker, B. S., Andonian, K., Swope, S. M., & Rieseberg, L. H. (2015). Evolution of invasiveness through increased resource use in a vacant niche. *Nature Plants*, 1(6), 1–5. <https://doi.org/10.1038/nplants.2015.66>

Dlugosch, K. M., Lai, Z., Bonin, A., Hierro, J., & Rieseberg, L. H. (2013). Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *Genes, Genomes, Genetics*, 3(2), 359–367. <https://doi.org/10.1534/g3.112.003871>

Doležel, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany*, 95(1), 99–110. <https://doi.org/10.1093/aob/mci005>

Doležel, J., Greilhuber, J., & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, 2(9), 2233–2244. <https://doi.org/10.1038/nprot.2007.310>

Ellis, L. L., Huang, W., Quinn, A. M., Ahuja, A., Alfrejd, B., Gomez, F. E., Hjelmen, C. E., Moore, K. L., Mackay, T. F. C., Johnston, J. S., & Tarone, A. M. (2014). Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. *PLoS Genetics*, 10(7), e1004522. <https://doi.org/10.1371/journal.pgen.1004522>

Eriksen, R. L., Hierro, J. L., Eren, Ö., Andonian, K., Török, K., Becerra, P. I., Montesinos, D., Khetsuriani, L., Diaconu, A., & Kesseli, R. (2014). Dispersal pathways and genetic differentiation among worldwide populations of the invasive weed *Centaurea solstitialis* L. (Asteraceae). *PLoS ONE*, 9(12), e114786. <https://doi.org/10.1371/journal.pone.0114786>

Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, 41(1), 331–368.  
<https://doi.org/10.1146/annurev.genet.40.110405.090448>

Francis, D., Davies, M. S., & Barlow, P. W. (2008). A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Annals of Botany*, 101(6), 747–757.  
<https://doi.org/10.1093/aob/mcn038>

Franks, P. J., Freckleton, R. P., Beaulieu, J. M., Leitch, I. J., & Beerling, D. J. (2012). Megacycles of atmospheric carbon dioxide concentration correlate with fossil plant genome size. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1588), 556–564. <https://doi.org/10.1098/rstb.2011.0269>

Fridley, J., & Craddock, A. (2015). Contrasting growth phenology of native and invasive forest shrubs mediated by genome size. *New Phytologist*, 207(3), 659–668.

Gaut, B. S., & Ross-Ibarra, J. (2008). Selection on major components of angiosperm genomes. *science*, 320(5875), 484-486.

Gerlach, J. D. (1997). How the west was lost: reconstructing the invasion dynamics of yellow starthistle and other plant invaders of western rangelands and natural areas. *Proceedings of the California Exotic Pest Plant Council Symposium*, 3, 67–72.

González, J., Lenkov, K., Lipatov, M., Macpherson, J. M., & Petrov, D. A. (2008). High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biology*, 6(10), e251. <https://doi.org/10.1371/journal.pbio.0060251>

González, J., Macpherson, J. M., Messer, P. W., & Petrov, D. A. (2008). Inferring the strength of selection in *Drosophila* under complex demographic models. *Molecular Biology and Evolution*, 26(3), 513–526. <https://doi.org/10.1093/molbev/msn270>

Goubert, C., Henri, H., Minard, G., Moro, C. V., Mavingui, P., Vieira, C., & Boulesteix, M. (2017). High-throughput sequencing of transposable element insertions suggests adaptive evolution of the invasive Asian tiger mosquito towards temperate environments. *Molecular Ecology*, 26(15), 3968–3981.  
<https://doi.org/10.1111/mec.14184>

Gralka, M., Stiewe, F., Farrell, F., Möbius, W., Waclaw, B., & Hallatschek, O. (2016). Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19(8), 889–898. <https://doi.org/10.1111/ele.12625>

Greilhuber, J. (2005). Intraspecific variation in genome size in angiosperms: identifying its existence. *Annals of Botany*, 95(1), 91–98. <https://doi.org/10.1093/aob/mci004>

Grotkopp, E., Rejmánek, M., Sanderson, M. J., & Rost, T. L. (2004). Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution*, 58(8), 1705–1729. <https://doi.org/10.1554/03-545>

Guignard, M. S., Nichols, R. A., Knell, R. J., Macdonald, A., Romila, C.-A., Trimmer, M., Leitch, I. J., & Leitch, A. R. (2016). Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytologist*, 210(4), 1195–1206. <https://doi.org/10.1111/nph.13881>

Heiser, C. B., & Whitaker, T. W. (1948). Chromosome number, polyploidy, and growth habit in California weeds. *American Journal of Botany*, 35(3), 179–186. JSTOR. <https://doi.org/10.2307/2438241>

Henry, R. C., Bartoń, K. A., & Travis, J. M. J. (2015). Mutation accumulation and the formation of range limits. *Biology Letters*, 11(1), 20140871. <https://doi.org/10.1098/rsbl.2014.0871>

Hodgson, J. G., Sharafi, M., Jalili, A., Díaz, S., Montserrat-Martí, G., Palmer, C., Cerabolini, B., Pierce, S., Hamzehee, B., Asri, Y., Jamzad, Z., Wilson, P., Raven, J. A., Band, S. R., Basconcelo, S., Bogard, A., Carter, G., Charles, M., Castro-Díez, P., ... Simmons, E. (2010). Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog? *Annals of Botany*, 105(4), 573–584. <https://doi.org/10.1093/aob/mcq011>

Irimia, R.-E., Montesinos, D., Eren, Ö., Lortie, C. J., French, K., Cavieres, L. A., Sotes, G. J., Hierro, J. L., Jorge, A., & Loureiro, J. (2017). Extensive analysis of native and non-native *Centaurea solstitialis* L. populations across the world shows no traces of polyploidization. *PeerJ*, 5, e3531. <https://doi.org/10.7717/peerj.3531>

Jian, Y., Xu, C., Guo, Z., Wang, S., Xu, Y., & Zou, C. (2017). Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-06153-8>

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., & Schulman, A. H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences*, 97(12), 6603–6607. <https://doi.org/10.1073/pnas.110587497>

Kang, M., Wang, J., & Huang, H. (2015). Nitrogen limitation as a driver of genome size evolution in a group of karst plants. *Scientific Reports*, 5(1), 1–8. <https://doi.org/10.1038/srep11636>

Kenton, A. Y., Rudall, P. J., & Johnson, A. R. (1986). Genome size variation in *Sisyrinchium* L. (Iridaceae) and its relationship to phenotype and habitat. *Botanical Gazette*, 147(3), 342–354. JSTOR.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23(3), 482–490. <https://doi.org/10.1093/molbev/msj057>

Knight, C. A. (2005). The large genome constraint hypothesis: evolution, ecology and phenotype. *Annals of Botany*, 95(1), 177–190. <https://doi.org/10.1093/aob/mci011>

Knight, C. A., & Beaulieu, J. M. (2008). Genome size scaling through phenotype space. *Annals of Botany*, 101(6), 759–766. <https://doi.org/10.1093/aob/mcm321>

Knight, Charles A., & Ackerly, D. D. (2002). Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters*, 5(1), 66–76. <https://doi.org/10.1046/j.1461-0248.2002.00283.x>

Kubešová, M., Moravcová, L., Suda, J., Jarošík, V., & Pyšek, P. (2010). Naturalized plants have smaller genomes than their non-invading relatives: A flow cytometric analysis of the Czech alien flora. *Preslia*, 82(1), 81–96.

Kuester, A., Conner, J. K., Culley, T., & Baucom, R. S. (2014). How weeds emerge: a taxonomic and trait-based examination using United States data. *New Phytologist*, 202(3), 1055–1068. <https://doi.org/10.1111/nph.12698>

Lertzman-Lepofsky, G., Mooers, A. Ø., & Greenberg, D. A. (2019). Ecological constraints associated with genome size across salamander lineages. *Proceedings of the Royal Society B: Biological Sciences*, 286(1911), 20191780. <https://doi.org/10.1098/rspb.2019.1780>

Lisch, D. (2013). How important are transposons for plant evolution? *Nature Reviews Genetics*, 14(1), 49–61. <https://doi.org/10.1038/nrg3374>

Lockton, S., Ross-Ibarra, J., & Gaut, B. S. (2008). Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences*, 105(37), 13965–13970. <https://doi.org/10.1073/pnas.0804671105>

Lockton, Steven, & Gaut, B. S. (2010). The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evolutionary Biology*, 10(1), 10. <https://doi.org/10.1186/1471-2148-10-10>

Lysák, M. (2000). Limited genome size variation in *Sesleria albicans*. *Annals of Botany*, 86(2), 399–403. <https://doi.org/10.1006/anbo.2000.1200>

Maddox, D. M., & Mayfield, A. (1985). Yellow starthistle infestations are on the increase. *California Agriculture*, 39(11), 10–12.

Martinez, M. A., Baack, E. J., Hovick, S. M., & Whitney, K. D. (2018). A reassessment of the genome size–invasiveness relationship in reed canarygrass (*Phalaris arundinacea*). *Annals of Botany*, 121(7), 1309–1318. <https://doi.org/10.1093/aob/mcy028>

Mei, W., Stetter, M. G., Gates, D. J., Stitzer, M. C., & Ross-Ibarra, J. (2017). Adaptation in plant genomes: Bigger is different. *bioRxiv*, 196501.

Meyerson, L. A., Pyšek, P., Lučanová, M., Wigginton, S., Tran, C.-T., & Cronin, J. T. (2020). Plant genome size influences stress tolerance of invasive and native plants via plasticity. *Ecosphere*, 11(5), e03145. <https://doi.org/10.1002/ecs2.3145>

Miskella, J. (2014). Hybridization between yellow starthistle (*Centaurea solstitialis*) and meadow knapweed (*Centaurea x moncktonii*) [MSc Thesis]. Oregon State University.

Münzbergová, Z. (2009). The effect of genome size on detailed species traits within closely related species of the same habitat. *Botanical Journal of the Linnean Society*, 160(3), 290–298. <https://doi.org/10.1111/j.1095-8339.2009.00973.x>

Noirot, M. (2005). Investigation on the Causes of Stoichiometric Error in Genome Size Estimation Using Heat Experiments: Consequences on Data Interpretation. *Annals of Botany*, 95(1), 111–118. <https://doi.org/10.1093/aob/mci006>

Noirot, M., Barre, P., Louarn, J., Duperray, C., & Hamon, S. (2000). Nucleus–Cytosol Interactions—A Source of Stoichiometric Error in Flow Cytometric Estimation of

Nuclear DNA Content in Plants. *Annals of Botany*, 86(2), 309–316.  
<https://doi.org/10.1006/anbo.2000.1187>

Ohri, D. (1998). Genome size variation and plant systematics. *Annals of Botany*, 82, 75–83.  
<https://doi.org/10.1006/anbo.1998.0765>

Oney-Birol, S., & Tabur, S. (2018). Genome size and morphological variations in *Brachypodium distachyon* L. along altitudinal levels. *Pakistan Journal of Botany*, 50(5), 1923–1933.

Öztürk, M., Martin, E., Dinç, M., Duran, A., Özdemir, A., & Çetin, Ö. (2009). A cytogenetical study on some plant taxa in Nizip region (Aksaray, Turkey). *Turkish Journal of Biology*, 33(1), 35–44.

Pandit, M. K., Pocock, M. J. O., & Kunin, W. E. (2011). Ploidy influences rarity and invasiveness in plants: Ploidy rarity and invasiveness. *Journal of Ecology*, 99(5), 1108–1115. <https://doi.org/10.1111/j.1365-2745.2011.01838.x>

Paule, J., Paule, L., & Gömöry, D. (2018). Small genome size variation across the range of European beech (*Fagus sylvatica*). *Plant Systematics and Evolution*, 304(4), 577–582.  
<https://doi.org/10.1007/s00606-018-1501-8>

Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22(24), 5972–5982.  
<https://doi.org/10.1111/mec.12524>

Petrov, D. A. (2003). Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Molecular Biology and Evolution*, 20(6), 880–892.  
<https://doi.org/10.1093/molbev/msg102>

Piegu, B., Guyot, R., Picault, N., Roulin, A., Saniyal, A., Kim, H., Collura, K., Brar, D. S., Jackson, S., Wing, R. A., & Panaud, O. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research*, 16(10), 1262–1269.  
<https://doi.org/10.1101/gr.5290206>

Pitcairn, M. J., Schoenig, S., Yacoub, R., & Gendron, J. (2006). Yellow starthistle continues its spread in California. *California Agriculture*, 60(2), 83–90.  
<https://doi.org/10.3733/ca.v060n02p83>

Price, H. (2000). Sunflower (*Helianthus annuus*) Leaves Contain Compounds that Reduce Nuclear Propidium Iodide Fluorescence. *Annals of Botany*, 86(5), 929–934.  
<https://doi.org/10.1006/anbo.2000.1255>

Prokopowich, C. D., Gregory, T. R., & Crease, T. J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46(1), 48–50.  
<https://doi.org/10.1139/g02-103>

Pyšek, P., Skálová, H., Čuda, J., Guo, W.-Y., Suda, J., Doležal, J., Kauzál, O., Lambertini, C., Lučanová, M., Mandáková, T., Moravcová, L., Pyšková, K., Brix, H., & Meyerson, L. A. (2018). Small genome separates native and invasive populations in an ecologically important cosmopolitan grass. *Ecology*, 99(1), 79–90. <https://doi.org/10.1002/ecy.2068>

Pyšek, P., Lučanová, M., Dawson, W., Essl, F., Kreft, H., Leitch, I. J., Lenzner, B., Meyerson, L.A., Pergl, J., van Kleunen, M., Weigelt, P., Winter, M., & Guo, W. Y. (2023). Small genome size and variation in ploidy levels support the naturalization of vascular plants but constrain their invasive spread. *New Phytologist*. <https://doi.org/10.1111/nph.19135>

Qiu, F., Baack, E. J., Whitney, K. D., Bock, D. G., Tetreault, H. M., Rieseberg, L. H., & Ungerer, M. C. (2019). Phylogenetic trends and environmental correlates of nuclear genome size variation in *Helianthus* sunflowers. *New Phytologist*, 221(3), 1609–1618. <https://doi.org/10.1111/nph.15465>

Reese, J. B., & Williams, J. H. (2019). How does genome size affect the evolution of pollen tube growth rate, a haploid performance trait? *American Journal of Botany*, 106(7), 1011–1020. <https://doi.org/10.1002/ajb2.1326>

Roddy, A. B., Thérioux-Rancourt, G., Abbo, T., Benedetti, J. W., Brodersen, C. R., Castro, M., Castro, S., Gilbride, A. B., Jensen, B., Jiang, G.-F., Perkins, J. A., Perkins, S. D., Loureiro, J., Syed, Z., Thompson, R. A., Kuebbing, S. E., & Simonin, K. A. (2020). The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *International Journal of Plant Sciences*, 181(1), 75–87. <https://doi.org/10.1086/706186>

Roessler, K., Muyle, A., Diez, C. M., Gaut, G. R., Bousios, A., Stitzer, M. C., Seymour, D. K., Doebley, J. F., Liu, Q., & Gaut, B. S. (2019). The genome-wide dynamics of purging during selfing in maize. *Nature plants*, 5(9), 980-990. <https://doi.org/10.1038/s41477-019-0508-7>

Schmuths, H. (2004). Genome size variation among accessions of *Arabidopsis thaliana*. *Annals of Botany*, 93(3), 317–321. <https://doi.org/10.1093/aob/mch037>

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschétzki, K., Weichselgartner, T., Kemen, C., Stökl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, 5(1), 1–10. <https://doi.org/10.1038/ncomms6495>

Slatkin, M., & Excoffier, L. (2012). Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, 191(1), 171–181. <https://doi.org/10.1534/genetics.112.139022>

Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*, 24(9), 2241–2252. <https://doi.org/10.1111/mec.13089>

Temsch, E. M., & Greilhuber, J. (2001). Genome size in *Arachis duranensis*: a critical study. *Genome*, 44(5), 826–830. <https://doi.org/10.1139/gen-44-5-826>

Temsch, E. M., Temsch, W., Ehrendorfer-Schratt, L., & Greilhuber, J. (2010). Heavy metal pollution, selection and genome size: the species of the Žerjav Study revisited with flow cytometry. *Journal of Botany*, 2010. <https://doi.org/10.1155/2010/596542>

Tenaillon, M. I., Hollister, J. D., & Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, 15(8), 471–478.  
<https://doi.org/10.1016/j.tplants.2010.05.003>

Tuna, G. S., Başer, İ., & Tuna, M. (2019). Genome size variation among natural populations of *Brachypodium distachyon* and *B. hybridum* collected from different regions of Turkey. *Turkish Journal of Botany*, 43(2), 196–207.

Veselý, P., Šmarda, P., Bureš, P., Stirton, C., Muasya, A. M., Mucina, L., Horová, L., Veselá, K., Šílerová, A., Šmerda, J., & Knápek, O. (2020). Environmental pressures on stomatal size may drive plant genome size evolution: evidence from a natural experiment with Cape geophytes. *Annals of Botany*, 126(2), 323–330.  
<https://doi.org/10.1093/aob/mcaa095>

Vielle-Calzada, J.-P., de la Vega, O. M., Hernández-Guzmán, G., Ibarra-Laclette, E., Alvarez-Mejía, C., Vega-Arreguín, J. C., Jiménez-Moraila, B., Fernández-Cortés, A., Corona-Armenta, G., Herrera-Estrella, L., & Herrera-Estrella, A. (2009). The Palomero genome suggests metal effects on domestication. *Science*, 326(5956), 1078–1078.  
<https://doi.org/10.1126/science.1178437>

Wam, H. K., Stolter, C., & Nybakken, L. (2017). Compositional change in foliage phenolics with plant age, a natural experiment in boreal forests. *Journal of Chemical Ecology*, 43(9), 920–928. <https://doi.org/10.1007/s10886-017-0881-5>

Whitney, K. D., Baack, E. J., Hamrick, J. L., Godt, M. J. W., Barringer, B. C., Bennett, M. D., ... & Ross-Ibarra, J. (2010). A role for nonadaptive processes in plant genome size evolution?. *Evolution: International Journal of Organic Evolution*, 64(7), 2097-2109.

Widmer, T. L., Guermache, F., Dolgovskaia, M. Y., & Reznik, S. Ya. (2007). Enhanced growth and seed properties in introduced vs. native populations of yellow starthistle (*Centaurea solstitialis*). *Weed Science*, 55(5), 465–473. <https://doi.org/10.1614/WS-06-211R.1>

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of mutational load at the edges of a species range. *Molecular Biology and Evolution*, 35(4), 781–791.  
<https://doi.org/10.1093/molbev/msy003>

Witzell, J., Gref, R., & Näsholm, T. (2003). Plant-part specific and temporal variation in phenolic compounds of boreal bilberry (*Vaccinium myrtillus*) plants. *Biochemical Systematics and Ecology*, 31(2), 115–127.  
[https://doi.org/10.1016/S0305-1978\(02\)00141-2](https://doi.org/10.1016/S0305-1978(02)00141-2)

Zhang, Q.-J., & Gao, L.-Z. (2017). Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3 (Bethesda, Md.)*, 7(6), 1875–1885. <https://doi.org/10.1534/g3.116.037572>

## Main Tables and Figures

**Table 1.** Field collections from the YST invaded range in California, USA.

			Latitude, Longitude	Elevation (m)	N (Common garden)	N (Genome size)
Coast	Diablo	DIA	N 37.863694, W 121.976444	237	20	21(1)
	Gilroy	GIL	N 37.03373, W 121.53674	68	22	27(2)
	Marin	MAR16	N 38.01154, W 122.612222	143	25	26(2)
	Napa	NAP	N 38.2400167, W 122.2702545	17	22	27(3)
Central Valley	Clovis	CLV	N 36.9231479, W 119.793704	118	17	18(2)
	Red Bluff	RB	N 40.27083, W 122.27104	157	23	24(1)
	Triangle	TRI	N 37.46178, W 119.79218	866	13	13
Sierra Nevadas	Arnold	ARN	N 38.262671, W 120.336713	1273	30	30(2)
	Belden	BEL	N 40.004779, W 121.261753	679	25	26(2)
	Lyons	LYO	N 38.078308, W 120.164527	1497	21	21(1)
	Placerville	PLA	N 38.733866, W 120.829686	507	21	23(2)
	Quincy	QUI	N 39.964903, W 120.944909	1075	23	24(2)
	Sandflat	SAN	N 38.76417, W 120.327177	1192	27	29(3)
	Vet	VET	N 38.09996, W 120.58947	386	23	24(1)
				Total	312	333(24)

Sample sizes of genome size estimates for each population include the overall number of individuals with at least 1 measurement. Values in parentheses indicate how many individuals out of the total have replicate estimates.

**Table 2.** Linear model explaining 2C genome size (pg) variation.

Effect type	Effect	Coefficient	df	F-ratio	<i>p</i>
Fixed	<b>population age</b>	<b>-0.00030</b>	<b>1</b>	<b>13.763</b>	<b>0.0002</b>
Fixed	<b>elevation</b>	<b>-0.00001</b>	<b>1</b>	<b>3.567</b>	<b>0.0598</b>
Fixed	<b>days until FC<sup>a</sup></b>	<b>-0.00031</b>	<b>1</b>	<b>13.022</b>	<b>0.0004</b>
Fixed	population age*elevation				NS
Fixed	population age*days until FC				NS
Fixed	elevation*days until FC				NS
Fixed	pop age*elevation*days until FC				NS
				$R^2$	0.0784
				$F_{(3,329)}$	9.324
				<i>p</i>	<0.0001

Significant effects are shown in bold. Effects without significant main or interaction effects (*p* >0.1) were removed from the model (NS).

<sup>a</sup>number of days between germination and date of flow cytometry estimation

**Table 3.** Best-fitting linear model explaining PC1 coordinates

Effect type	Effect	Coefficient	df	F-ratio	<i>p</i>
Fixed	genome size <sup>a</sup>	-6.752	1	-0.857	0.392
Fixed	<b>greenhouse position<sup>b</sup></b>	<b>see SI<sup>c</sup></b>	<b>8</b>	<b>3.411</b>	<b>0.0009</b>
Fixed	<b>genome size*greenhouse position</b>	<b>see SI<sup>c</sup></b>	<b>8</b>	<b>2.008</b>	<b>0.0454</b>
Fixed	days to harvest				NS
Fixed	genome size*days to harvest				NS
Fixed	greenhouse position*days to harvest				NS
Fixed	higher order interactions				NS
Random	block				NS
		$R^2$			0.126
		$F_{(17,294)}$			2.481
		<i>p</i>			0.001

PC1 coordinate values represent a composite of growth rate and biomass traits. Significant effects are in bold. Effects without significant main or interaction effects (*p* > 0.1) were removed from the model (NS).

<sup>a</sup> 2C genome size corrected for effect of estimation date

<sup>b</sup> distance to cooling pads in greenhouse, scored as positions 1 (furthest) - 9 (closest)

<sup>c</sup> see Supplementary Table 1 for all level and interaction coefficients

**Table 4.** Best-fitting linear model explaining PC2

Effect type	Effect	Coefficient	df	F-ratio	<i>p</i>
Fixed	<b>genome size<sup>b</sup></b>	<b>6.476</b>	<b>1</b>	<b>3.209</b>	<b>0.001</b>
Fixed	<b>days to harvest</b>	<b>-0.034</b>	<b>1</b>	<b>3.63</b>	<b>0.058</b>
Fixed	<b>greenhouse position<sup>b</sup></b>	<b>see SI<sup>c</sup></b>	<b>8</b>	<b>2.09</b>	<b>0.037</b>
Fixed	<b>greenhouse position*days to harvest</b>	<b>see SI<sup>c</sup></b>	<b>8</b>	<b>2.22</b>	<b>0.026</b>
Fixed	genome size*days to harvest				NS
Fixed	genome size*greenhouse position				NS
Fixed	genome size*position*days to harvest				NS
Random	block				NS
		$R^2$			0.139
		$F_{(18,293)}$			2.631
		<i>p</i>			<0.001

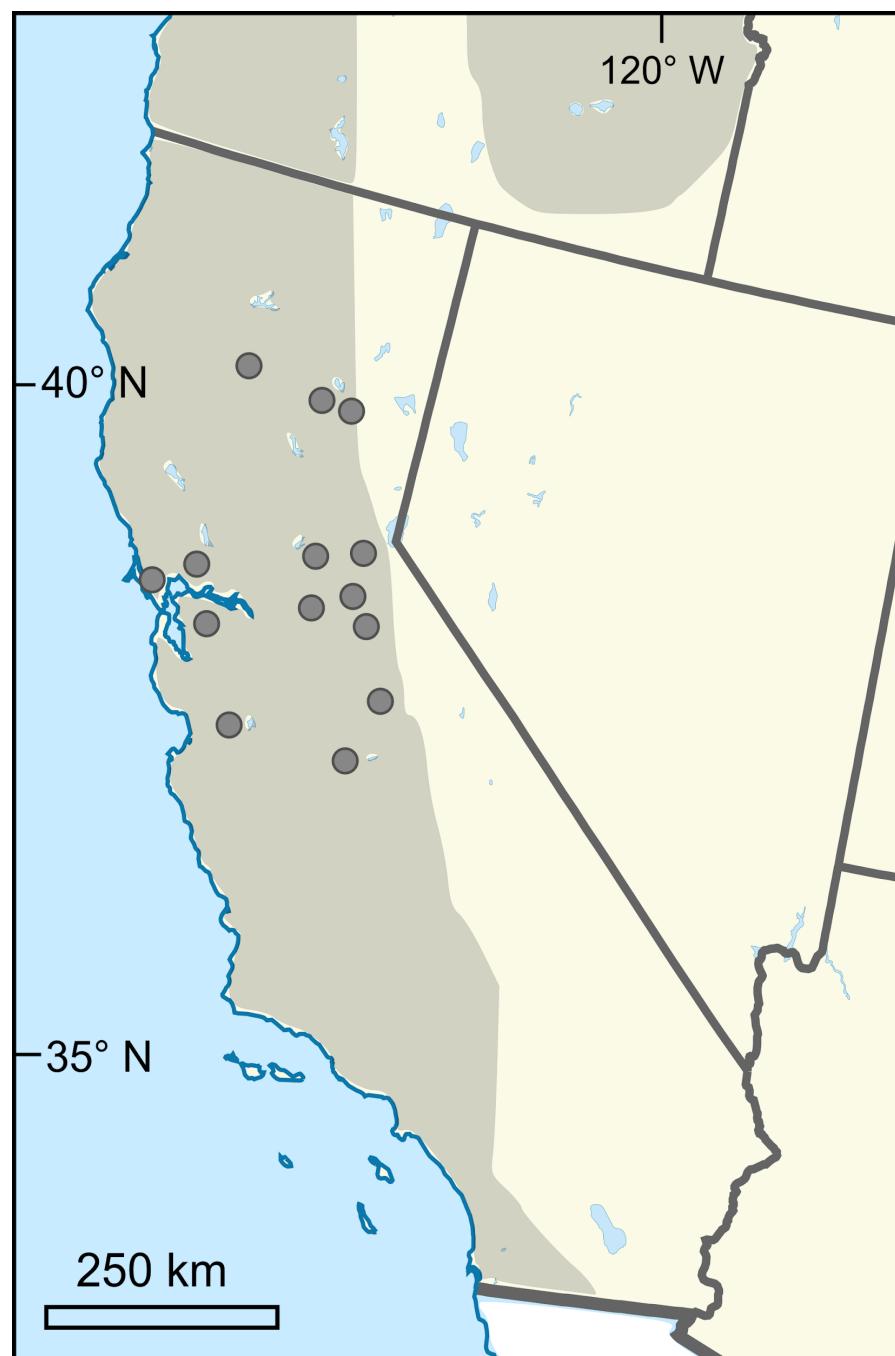
PC2 coordinate values represent a composite of development time and flowering traits.

Significant effects are shown in bold. Effects without significant main or interaction effects (*p* >0.1) were removed from the model (NS).

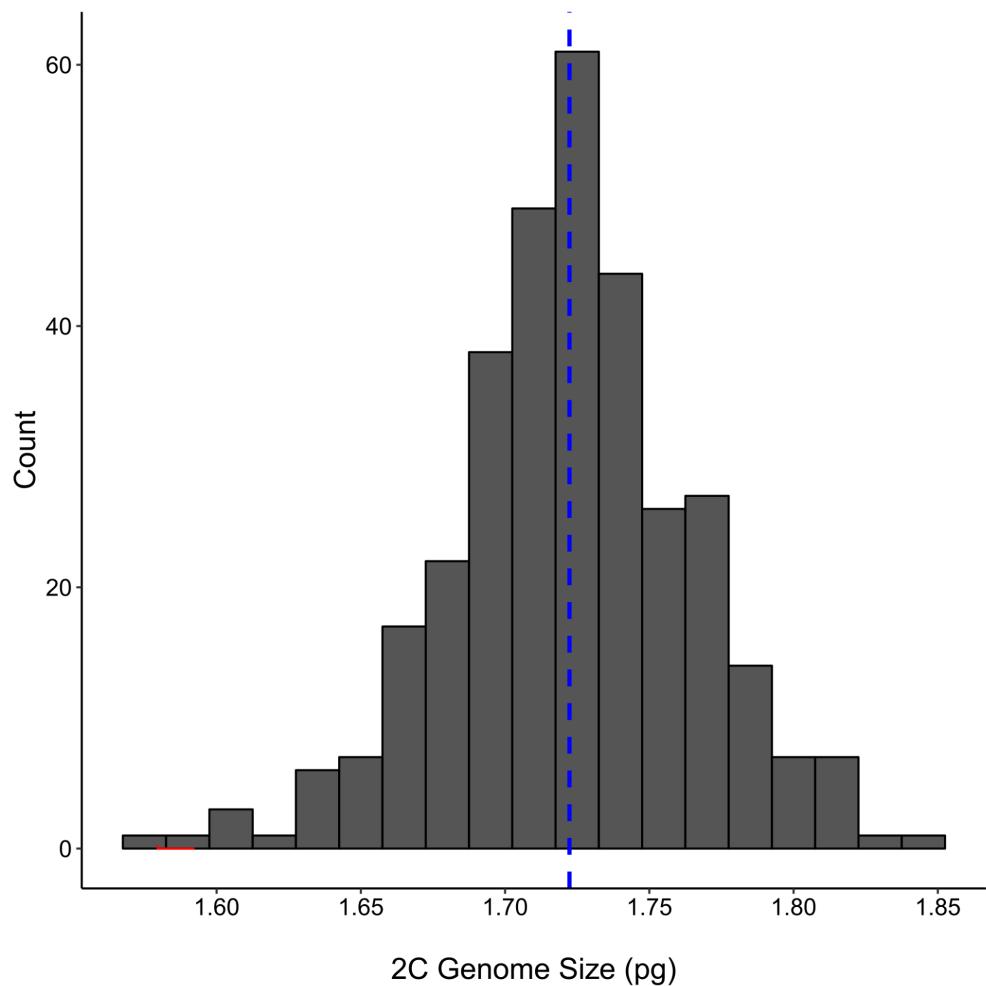
<sup>a</sup> 2C genome size corrected for effect of estimation date

<sup>b</sup> distance to cooling pads in greenhouse, scored as positions 1 (furthest) - 9 (closest)

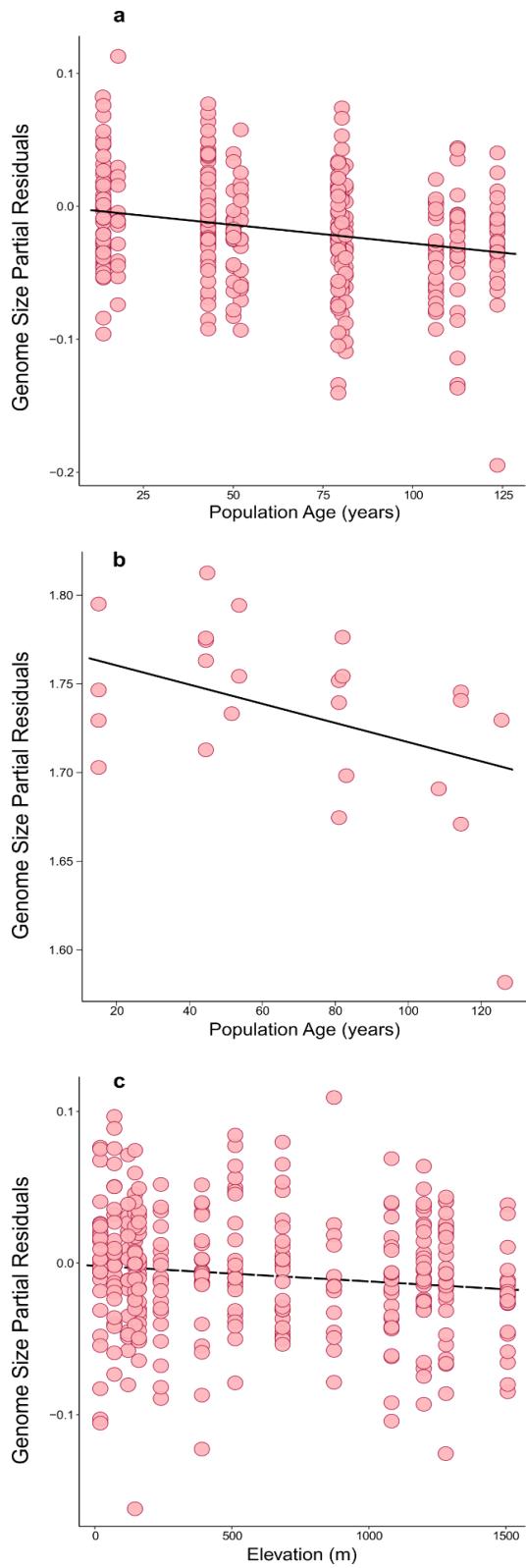
<sup>c</sup> see Supplementary Table 2 for all level and interaction coefficients



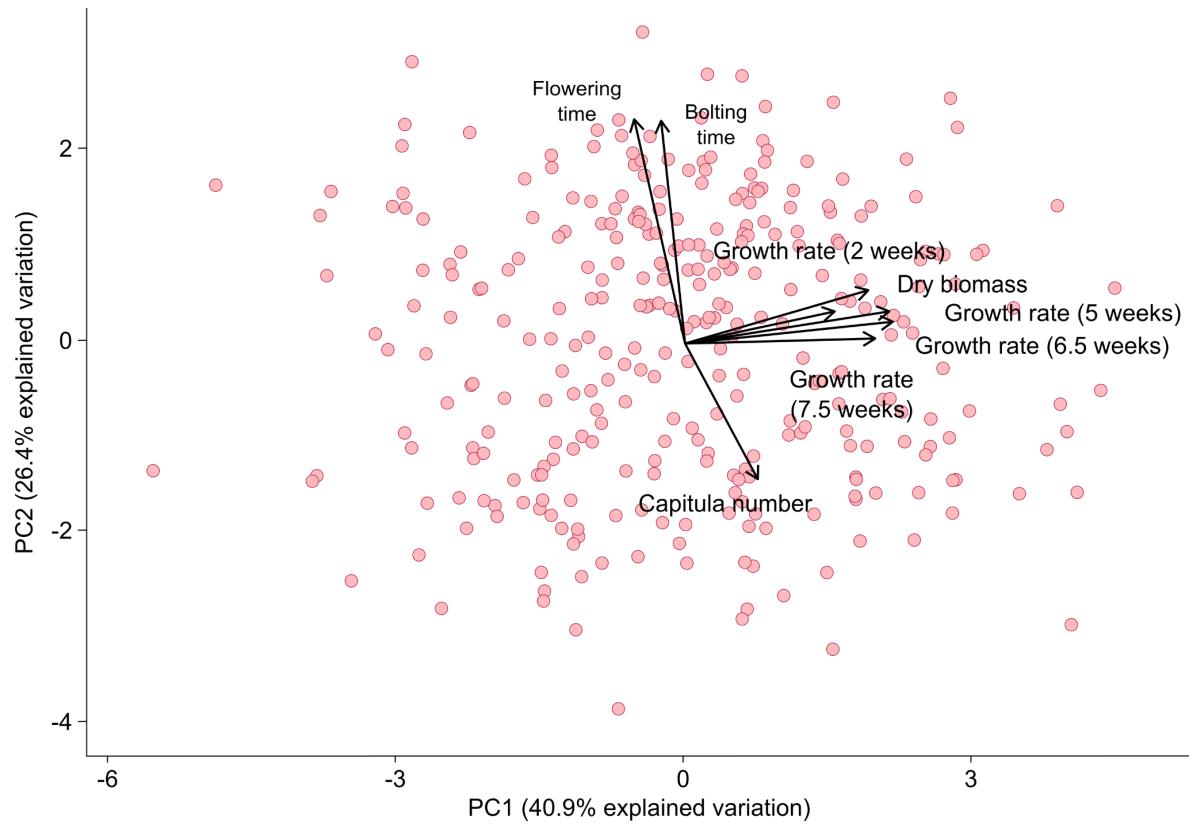
**Figure 1**



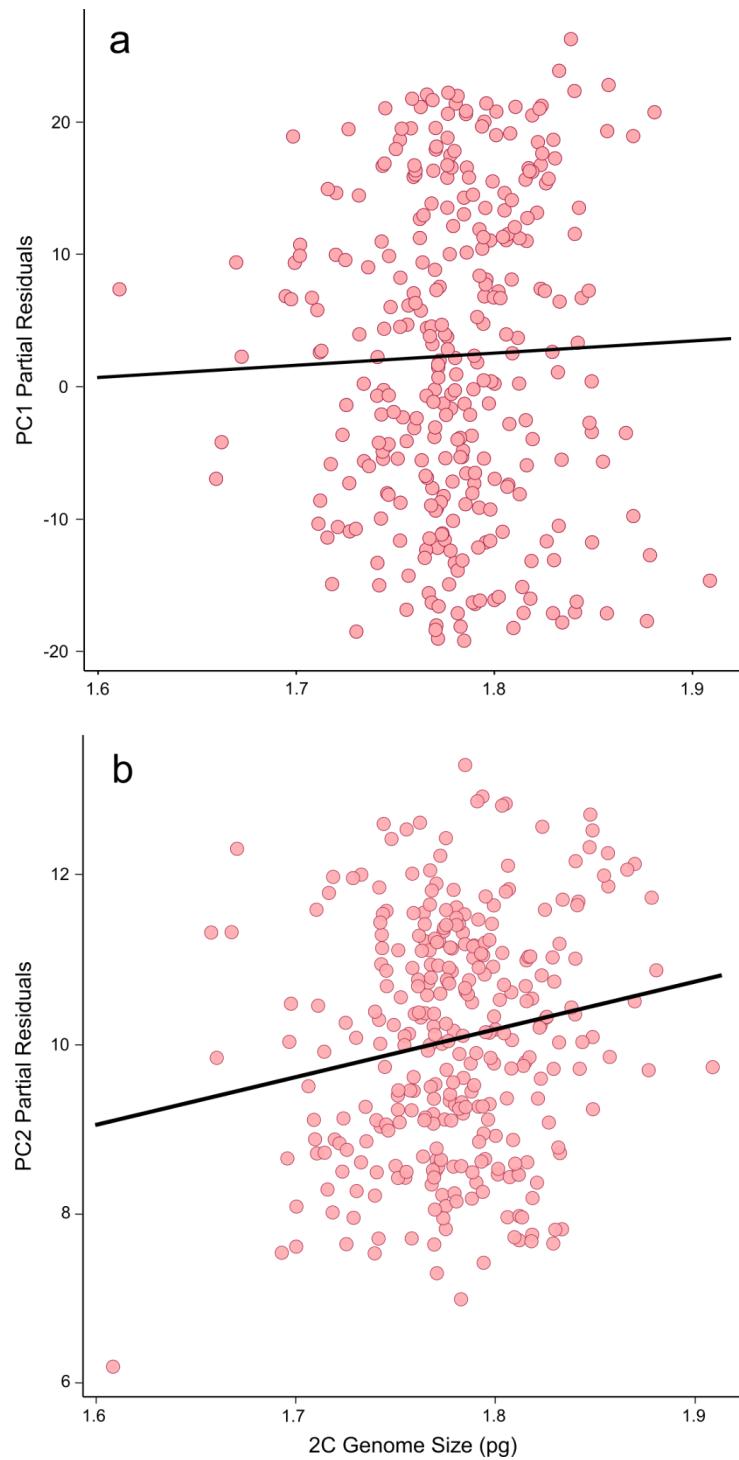
**Figure 2**



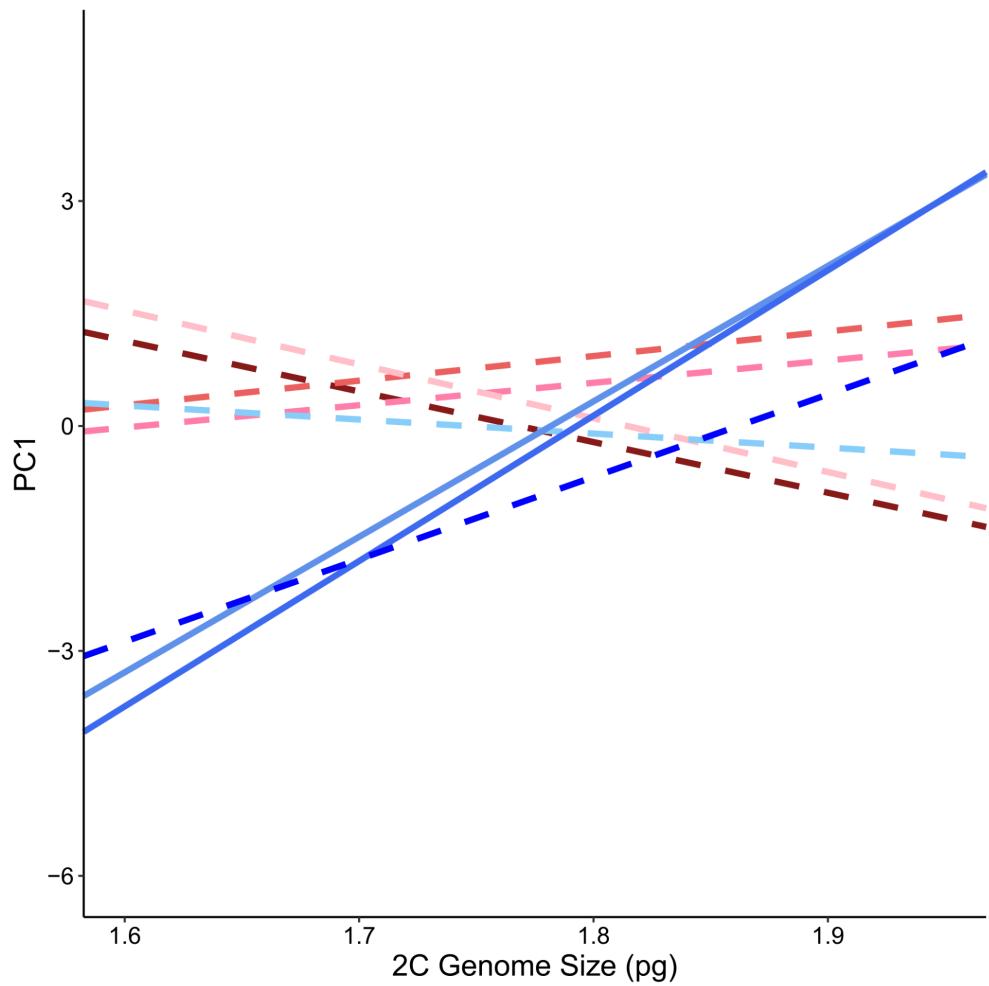
**Figure 3**



**Figure 4**



**Figure 5**



**Figure 6**

**Figure 1.** Map of 14 YST populations sampled in 2016 from the invasion in California, USA.

Shaded areas indicate the invaded range in western North America.

**Figure 2.** Histogram of 2C genome sizes sampled from the Californian invasion (N=335), with a mean size of 1.72pg (blue dashed line). For individuals with repeat measurements (N=24), only the means of estimated genome sizes are included.

**Figure 3.** Partial residuals from best-fitting linear models explaining patterns of genome size variation using the full genome size data set of California individuals (N=335) by (a) population age and (c) elevation. (b) Partial residuals by population age using a reduced dataset of only individuals with repeat flow cytometry measurements (N=24). Solid lines indicate significant associations ( $p<0.05$ ) and a dashed line indicates a marginally significant association ( $p<0.1$ ).

**Figure 4.** Principal component analysis of traits where each point indicates an individual plant in the common garden (N=391) with PC1 and PC2 explaining 67.3% of all variation. PC1 is associated with variation in growth traits and PC2 is associated with variation in development and

**Figure 5.** (a) Partial residual plot from best-fitting model explaining variation in PC1 scores due to fixed effect of genome size (NS,  $p=0.39$ ). (b) Partial residual plot explaining variation in PC2 scores by genome size ( $p=0.001$ )

**Figure 6.** Genome size varies in direction and magnitude of effect on PC1 scores when plotted by greenhouse block. Different colour lines from red to blue indicate greenhouse position with increasing proximity to the cooling pads, where solid lines indicate a significant effect of genome

size on PC1 ( $p<0.05$ ). Significant and marginally significant positive correlations were found in positions 7 and 8, and position 9, respectively, where local environments were coolest.