

Repeatability of flowering phenology in *Echinacea angustifolia*

(Brief) Introduction

The narrow leaf purple coneflower, *Echinacea angustifolia*, is a long-lived plant native to northern tall-grass prairie remnants. *E.angustifolia* is a self-incompatible species, therefore successful reproduction relies both on available pollinators and pollen. Previous research has suggested that flowering earlier is selectively favored, both in *E. angustifolia* (Ison and Wagenius 2014; Ison et al. 2018) and across other plant species (Munguía-Rosas et al. 2011). While we know that *E.angustifolia* individuals that flower early have, on average, higher seed set than later flowering individuals (Ison and Wagenius 2014), we don't know if individuals are consistent in their flowering phenologies. This study aims to address the repeatability of phenological traits in *E.angustifolia*.

Here we are defining repeatability as the variation attributable to between subject differences, rather than within subject variation (which can be thought of as the phenotypic plasticity) (Nakagawa and Schielzeth 2010). In other words, repeatability is the within group correlation among measurements, where group can represent various meaningful forms (individuals, populations, species, etc) (Stoffel, Nakagawa, and Schielzeth 2017). Formally, repeatability can be quantified as:

$$R = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

where the numerator is the between group variance, and the denominator is the total phenotypic variation (Nakagawa and Schielzeth 2010). It's worth emphasizing that measurements of repeatability using this formulation can be low due to high within group variation or low between group variation (Nakagawa and Schielzeth 2010).

Study Questions

1. Is flowering phenology (i.e. timing) a repeatable trait in the perennial prairie plant, *Echinacea angustifolia*?
2. Is flowering duration a repeatable trait in *E.angustifolia*?
3. Does the repeatability of either of these traits depend on the origin site of these plants?

Methods

We used a long-term common garden experimental plot that was planted with *E.angustifolia* seeds from 8 different remnant sites in 1996. Remnants vary in size: the largest is a 45 ha preserve with 1000-6000 flowering plant per year owned by the Nature Conservancy and the smallest is less than one ha with 0-4 flowering plants per year. Individuals in the experimental plot are planted on a grid and the plot is similar to nearby populations in topography, density of plants, and pollinator community. *E.angustifolia* plants typically flower from late June to early August. Individuals don't usually begin flowering until their third year, and most plants don't flower every year, instead persisting as basal rosettes (Wagenius 2004). When they do flower, plants send up at least one (and up to 10) flowering head(s) (capitula). Each flowering head consists of row of sterile ray florets at the base and a series of disk florets arranged in concentric rings up the conical head. All the anthers in a single row of disc florets shed pollen on the same day and flowering progresses from the bottom towards the top of the conical head.

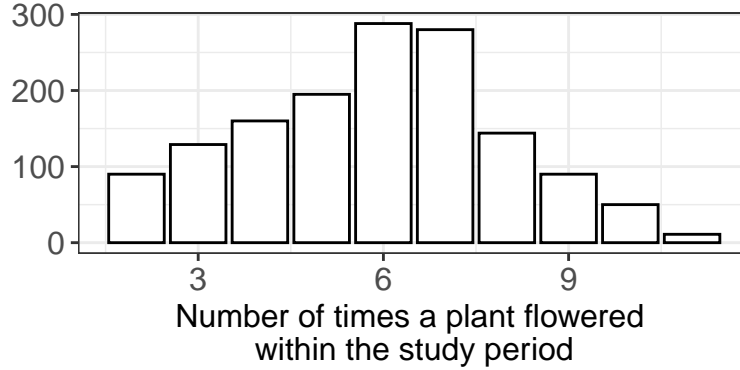


Figure 1: Figure 1. Number of times plants flowered during the study period. Plants that flowered zero or one time were excluded from the analysis

Flowering phenology data has been collected in the common garden experiment from 2005 to 2019. During the flowering season plants are checked every other day. A plant is recorded as having begun flowering when its anthers begin shedding pollen. Plants were monitored until they ended flowering.

Plants that did not flower ($n = 321$) or only flowered once ($n = 36$) during the study period were excluded, which left 289 plants in the analysis. Of those 289 plants, the average number of times an individual flowered was 5.88 (Figure 1). If a plant had multiple heads, we used the earliest start date and the latest end date as the full flowering period. We used the first flowering day (FFD) as a proxy for flowering time. The FFD is defined as the first day a flowering head sheds pollen. FFD has been used as measure of flowering phenology both in *E. angustifolia* (Ison and Wagenius 2014) and in other prairie species (Craine et al. 2012).

We used the R package rptR to analyze the repeatability of flowering phenology (Stoffel, Nakagawa, and Schielzeth 2017). P-values are determined via likelihood ratio tests that compare the fitted model to a model where the grouping variable (here most often cgPlaId) is removed (Stoffel, Nakagawa, and Schielzeth 2017). Since year-to-year fluctuations in weather can affect overall timing of flowering phenology (Figure 2), year was included as a fixed effect in most models.

Results using rptR package

Repeatability of FFD: Repeatability has been adjusted for year due to the year to year fluctuations (Figure 2). To adjust for year, we've included year as a fixed effect in our mixed model (Stoffel, Nakagawa, and Schielzeth 2017).

```
##
## Repeatability estimation using the lmm method
##
## Call = rpt(formula = startNum ~ year + (1 | cgPlaId), grname = "cgPlaId", data = phen_1996, datatype = "lmm")
##
## Data: 1437 observations
## -----
##
## cgPlaId (289 groups)
##
## Repeatability estimation overview:
##      R      SE  2.5% 97.5% P_permut  LRT_P
##  0.237 0.0263 0.192 0.296      0.01    0
##
## Bootstrapping and Permutation test:
```

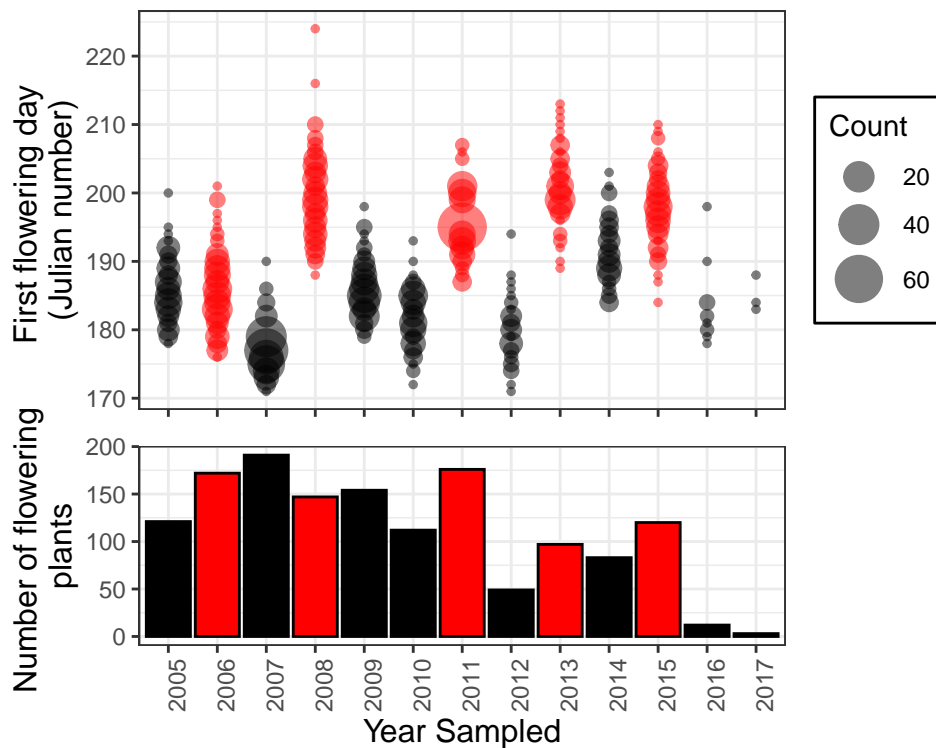
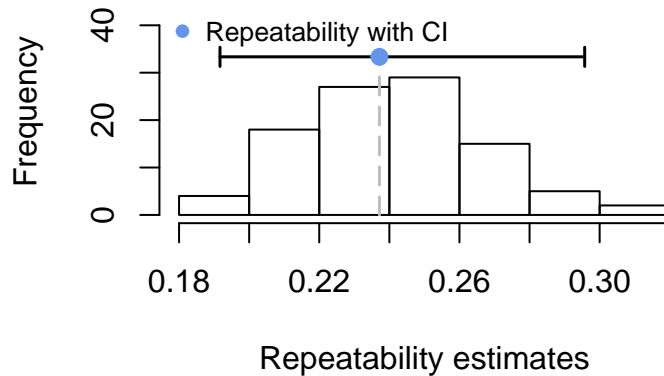


Figure 2: Figure 2. Number of flowering plants per year. Years when the experiment was burned are in red

```
##           N   Mean   Median   2.5%  97.5%
## boot      100 0.2405 0.240108 0.192 0.2957
## permut    100 0.0112 0.000147 0.000 0.0378
##
## Likelihood ratio test:
## logLik full model = -4029.197
## logLik red. model = -4082.045
## D = 106, df = 1, P = 4.3e-25
##
## -----
```

Bootstrap repeatabilities for cgPlaId



After adjusting for year, our estimate of the repeatability of FFDs for individuals is 0.237. What I think this means is that nearly 24% of the variation we see in FFDs, after accounting for yearly differences, is due to differences between individuals and 76% is due to variation within individuals. (Note: we can run the same analysis on FFDs that have been standardized by the median FFD of each year, and therefore remove the “year” adjustment from the model. This yields a similar estimate of repeatability ($R = 0.236$)).

Repeatability of FFDs by site: Instead of having the grouping variable of interest be the cgPlaId, I’ve changed it to site and re-run the model. Perhaps I need to be structuring the model differently, because I can’t figure out why this continues to be a singular fit... I’ve also included a model where I fit by *both* cgPlaId and site as grouping variables. Again this is a singular fit, but it makes more sense for me to fit the model like this, where cgPlaId is nested within site, so to speak.

Just with site only:

```
##
## Repeatability estimation using the lmm method
##
## Call = rpt(formula = startNum ~ year + (1 | site), grname = "site", data = phen_1996, datatype = "Ga
##
## Data: 1437 observations
## -----
##
## site (7 groups)
##
## Repeatability estimation overview:
##      R      SE    2.5% 97.5% P_permut  LRT_P
##  0.018  0.011 0.00101 0.0474    0.01     0
##
## Bootstrapping and Permutation test:
##      N    Mean Median    2.5% 97.5%
## boot   100 0.01664 0.0163 0.00101 0.04738
## permut  100 0.00138 0.0000 0.00000 0.00814
##
## Likelihood ratio test:
## logLik full model = -4073.767
## logLik red. model = -4082.045
## D = 16.6, df = 1, P = 2.36e-05
```

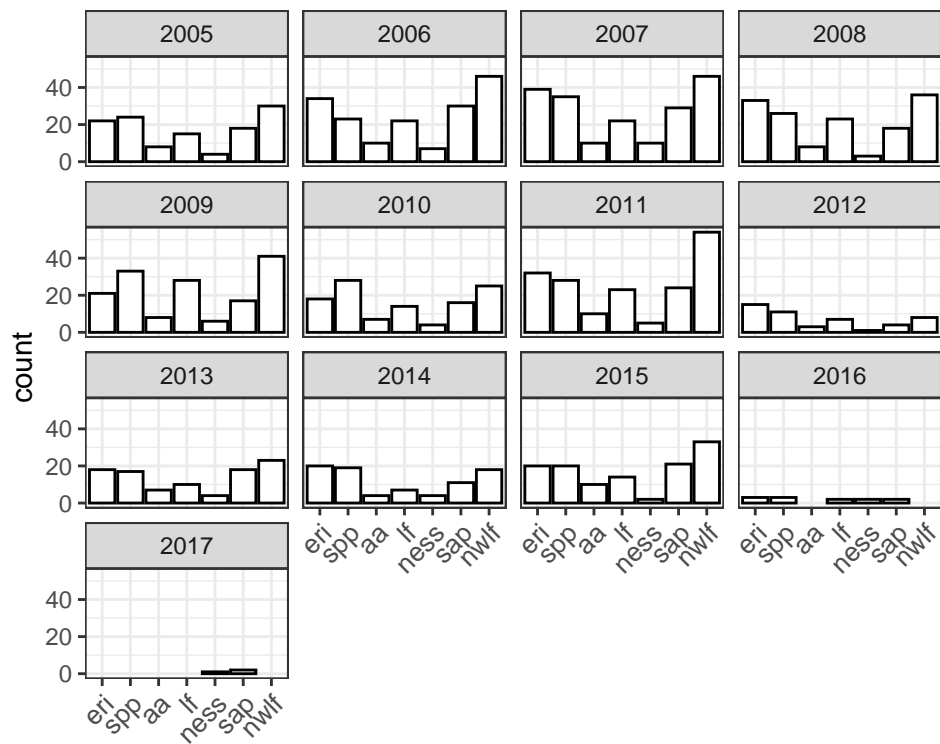
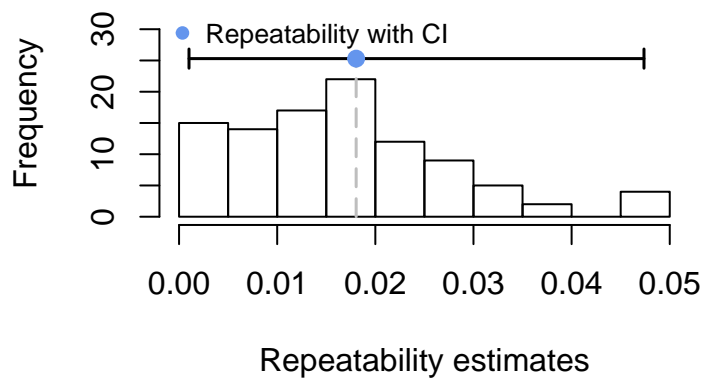


Figure 3: Figure 3. Number of flowering plants by site per year

##

Bootstrap repeatabilities for site



With site and cgPlaId as grouping variables

```
r3 <- rpt(startNum ~ year + (1|cgPlaId) + (1|site), grname = c("cgPlaId", "site"), data = phen_1996, da
summary(r3)
```

##

```
## Repeatability estimation using the lmm method
##
## Call = rpt(formula = startNum ~ year + (1 | cgPlaId) + (1 | site), grname = c("cgPlaId", "site"), da
##
## Data: 1437 observations
## -----
##
## cgPlaId (289 groups)
##
## Repeatability estimation overview:
##      R      SE   2.5% 97.5% P_permut  LRT_P
## 0.219 0.0294 0.164 0.276   0.002     0
##
## Bootstrapping and Permutation test:
##      N      Mean  Median   2.5% 97.5%
## boot   500 0.21940 2.21e-01 0.164 0.2757
## permut  500 0.00756 8.48e-07 0.000 0.0357
##
## Likelihood ratio test:
## logLik full model = -4024.51
## logLik red. model = -4070.817
## D = 92.6, df = 1, P = 3.18e-22
##
## -----
##
##
## site (7 groups)
##
## Repeatability estimation overview:
##      R      SE   2.5% 97.5% P_permut  LRT_P
## 0.0174 0.016     0 0.0598   0.07 0.006
##
## Bootstrapping and Permutation test:
##      N      Mean  Median   2.5% 97.5%
## boot   500 0.01725 0.01337 0.00e+00 0.0598
## permut  500 0.00865 0.00802 4.76e-14 0.0201
##
## Likelihood ratio test:
## logLik full model = -4024.51
## logLik red. model = -4027.62
## D = 6.22, df = 1, P = 0.00632
##
## -----
plot(r3)
```

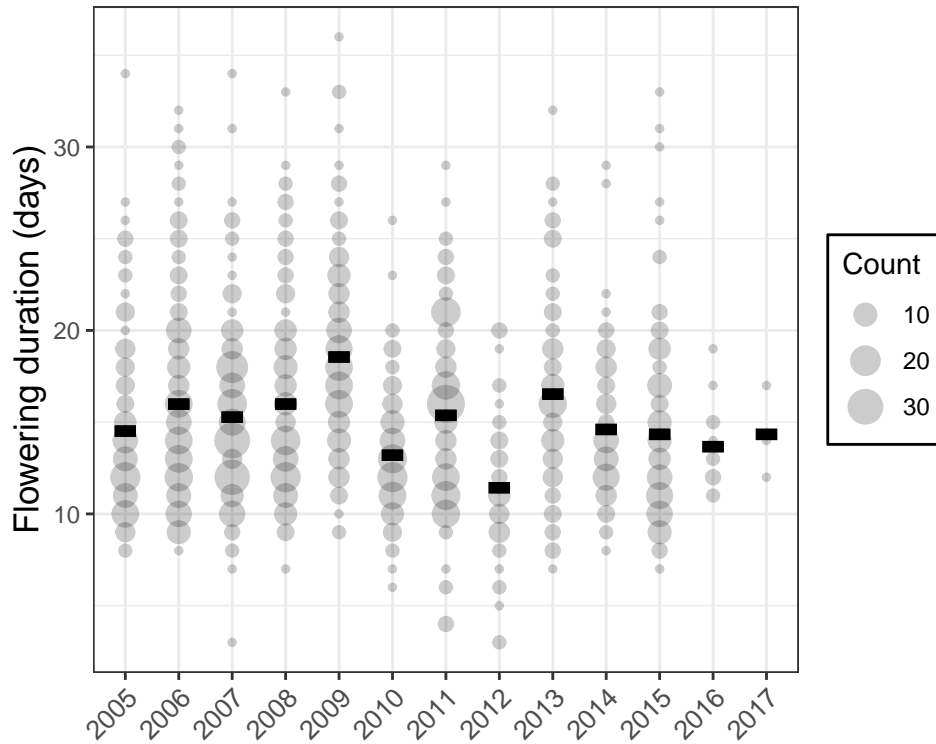
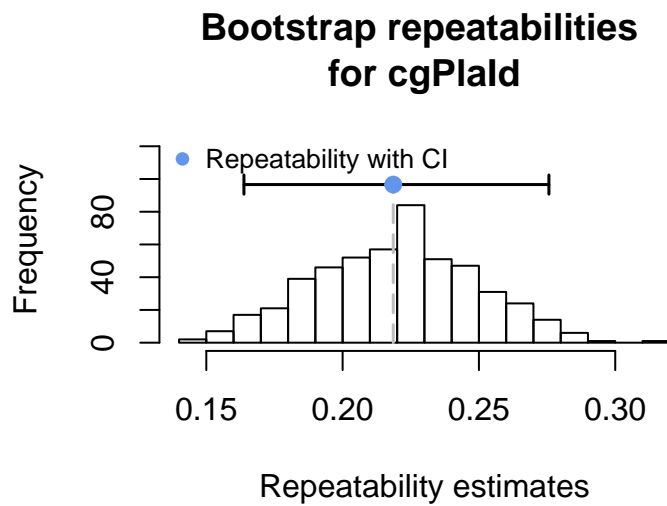


Figure 4: Figure 4. Length of flowering time by year. Average flowering time for each year shown as black bar

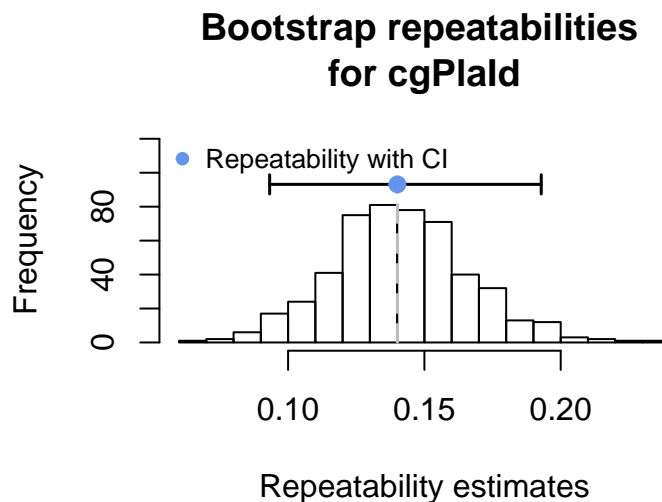


Repeatability of flowering duration: The length of flowering time is repeatable with and without adjusting for year. However, a simple mixed model comparison suggest that a model with year included provides a better fit than one without.

Output without adjusting for year:

```
##
## Repeatability estimation using the lmm method
```

```
##
## Call = rpt(formula = dur ~ (1 | cgPlaId), grname = "cgPlaId", data = phen_1996, datatype = "Gaussian"
##
## Data: 1437 observations
## -----
##
## cgPlaId (289 groups)
##
## Repeatability estimation overview:
##      R      SE   2.5% 97.5% P_permut  LRT_P
##    0.14 0.0255 0.0932 0.193   0.002    0
##
## Bootstrapping and Permutation test:
##      N      Mean   Median   2.5% 97.5%
## boot   500 0.14120 1.41e-01 0.0932 0.1929
## permut  500 0.00682 2.43e-09 0.0000 0.0331
##
## Likelihood ratio test:
## logLik full model = -4282.143
## logLik red. model = -4306.724
## D = 49.2, df = 1, P = 1.18e-12
##
## -----
```

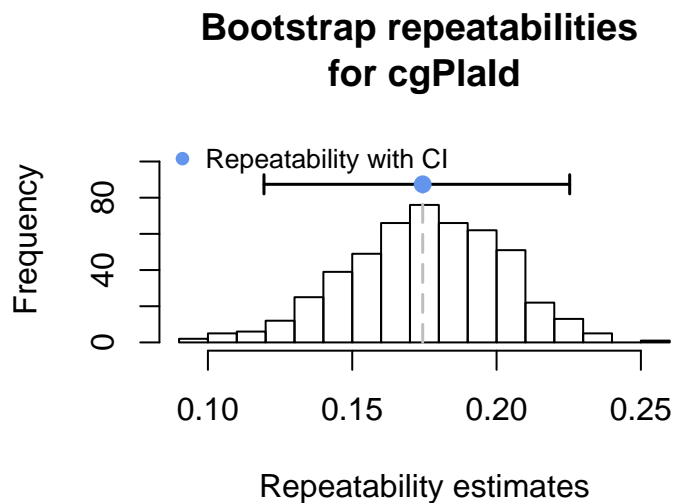


Output after adjusting for year:

```
##
## Repeatability estimation using the lmm method
##
## Call = rpt(formula = dur ~ year + (1 | cgPlaId), grname = "cgPlaId", data = phen_1996, datatype = "Gaussian"
##
## Data: 1437 observations
## -----
##
## cgPlaId (289 groups)
##
```



```
## Repeatability estimation overview:
##      R      SE   2.5% 97.5% P_permut  LRT_P
## 0.174 0.0264 0.119 0.225 0.002 0
##
## Bootstrapping and Permutation test:
##      N      Mean  Median   2.5% 97.5%
## boot   500 0.17523 0.17566 0.119 0.2253
## permut 500 0.00872 0.00282 0.000 0.0367
##
## Likelihood ratio test:
## logLik full model = -4194.439
## logLik red. model = -4230.275
## D = 71.7, df = 1, P = 1.27e-17
##
## -----
```



Similar to FFD, including site as a grouping variable results in singular fit models and does not appear to have a strong signal in terms of repeatability of flowering duration.

Results using classic correlations

Repeatability/consistency can also be quantified more simply as the correlation between measurements in any given year. Here I've just included a heatmap of the Spearman rank correlation between FFDs standardized by median FFD. As you can see, some years are better correlated with others, even after standardizing dates. However most years are at least marginally positively correlated with each other, further supporting the idea that flowering phenology is a consistent trait.

(Brief) Conclusions

Using the mixed model analysis framework, our results show that both the timing and the duration of flowering in *E. angustifolia* individuals are repeatable traits. Timing seems to be a more strongly consistent trait than duration. Our results also suggest that origin site is not meaningful in quantifying repeatability. Whether this is due to high variability within sites, or low variability between sites is, as of right now, uncertain.

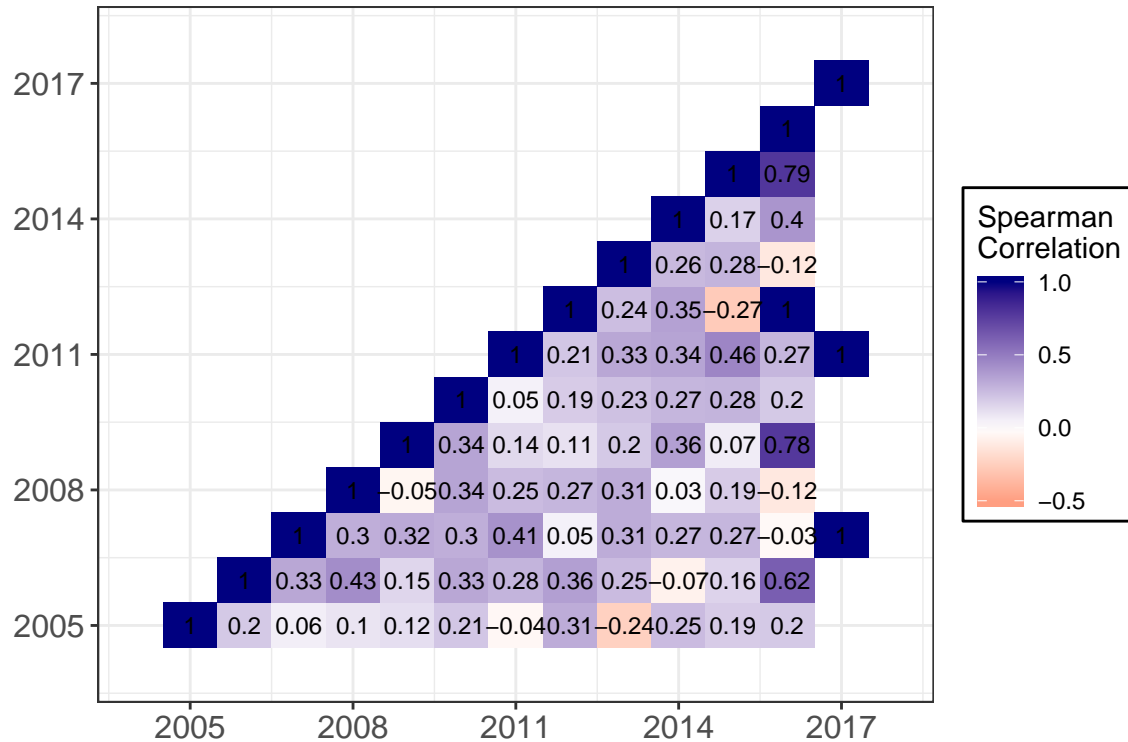


Figure 5: Figure 5. Spearman rank correlation heatmap of standardized FFDs. Only pairwise complete cases were used for each correlation.

Additional notes/questions:

- Right now we have shown that flowering phenology is a repeatable trait, but we haven't shown whether it's environmental vs. genetic. To tease this apart, it might be beneficial to do some sort of analysis of position in the common garden (I had included this in my previous analyses). If repeatability is spatially auto-correlated, then that might suggest that it's an environmental trait. If it's not, then I think we have stronger evidence that it's genetic.
- It'll also probably be necessary to quantify overlap between early and late flowering individuals so that we can address whether there is reproductive isolation in the population.
- Should I repeat the analysis for the 97, 98, 99 gardens? See how they compare?

References

- Craine, Joseph M., Elizabeth M. Wolkovich, E. Gene Towne, and Steven W. Kembel. 2012. "Flowering phenology as a functional trait in a tallgrass prairie." *New Phytologist*. <https://doi.org/10.1111/j.1469-8137.2011.03953.x>.
- Ison, Jennifer L., Leah J. Prescott, Scott W. Nordstrom, Amy Waananen, and Stuart Wagenius. 2018. "Pollinator-mediated mechanisms for increased reproductive success in early flowering plants." *Oikos*. <https://doi.org/10.1111/oik.04882>.
- Ison, Jennifer L., and Stuart Wagenius. 2014. "Both flowering time and distance to conspecific plants affect reproduction in *Echinacea angustifolia*, a common prairie perennial." *Journal of Ecology*. <https://doi.org/10.1111/1365-2745.12262>.

- Munguía-Rosas, Miguel A., Jeff Ollerton, Victor Parra-Tabla, and J. Arturo De-Nova. 2011. “Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured.” <https://doi.org/10.1111/j.1461-0248.2011.01601.x>.
- Nakagawa, Shinichi, and Holger Schielzeth. 2010. “Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists.” <https://doi.org/10.1111/j.1469-185X.2010.00141.x>.
- Stoffel, Martin A., Shinichi Nakagawa, and Holger Schielzeth. 2017. “rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models.” <https://doi.org/10.1111/2041-210X.12797>.
- Wagenius, Stuart. 2004. “Style persistence, pollen limitation, and seed set in the common prairie plant *Echinacea angustifolia* (Asteraceae).” *International Journal of Plant Sciences*. <https://doi.org/10.1086/386376>.