Flowering Schedules of *Echinacea angustifolia*

Rachael Sarette, Northwestern University

INTRODUCTION

Most of the prairie in the Midwest has been destroyed and what is left behind are small, scattered prairie remnants. Habituating these remnants are *Echinacea angustifolia*, a model prairie plant, that is self-incompatible. Pollen limitation is a problem facing *Echinacea* living in prairie remnants as there is reduced pollinator visitations and reduced compatible pollen transfer (Ison & Wagenius 2014), so timing of flowering and pollination are important for reproductive success and plants maximize the probability of being pollinated when they align their flowering schedules with those of nearby *Echinacea*.

One way to model the flowering schedule is by fitting the flowering data to an exponential sine function of the form $a*\sin[\pi((t-b)/c)^d]^e$ where the variables act as direct surrogates to relevant flowering features like maximum, start date, duration, lateness and length of tails (Malo 2002). This model accounts for over 90% of the data variance from unimodal flowers and allows us to compare these parameters across different *Echinacea* populations (Malo 2002).

In this study we investigated the extent to which flowering schedules differ between plants with only one head, called a solo head, and plants with multiple heads, called ensemble heads. We also investigated the extent to which the flowering schedules differ by population when grown in a common environment. We also looked at how measures of synchrony vary between three different perspectives: number of plants blooming at once, number of florets open at once, and floret estimates from the sine function.

METHODS

Data collection

The data utilized in this analysis was collected in 2005 with details in Ison & Wagenius (2014). Then we fit the exponential sine function to the data and determined the value of the 5 equation parameters.

Solo vs. Ensemble heads

Figure 1 shows the distribution of the values that the 5 parameters are assigned. Bootstrap analysis was conducted on the five parameters that determine the shape of the flowering schedule to see if the variables differed between solo and ensemble heads. Then we used a MANOVA test to determine the potential effects of uniqueness on all five variables.

Population differences

First a MANOVA test was run to see which variables are important in affecting the flowering schedules and thus define a population, then follow up univariate tests were run to determine how the flowering schedules differ between populations. Then we utilized the raw data for heads and combined it by plant to get a total achene count per day per plant. Then exponential sine functions were fit to this data to see the flowering schedules for each plant. This method worked for 210 of the plants, with only two plants not being able to be modeled by this method. These to plants had to be thrown out since there was no overlap in the timing of flowering between the heads on the plant so the plant did not have a unimodal flowering schedule which violated the assumptions of the model. The graphs of these heads can be seen in figure 2.

Comparing Perspectives

Flowering schedules calculated using these three perspectives used to compare synchrony measures are the plant view, floret view and the population schedule. The plant view was calculated by determining the number of heads flowering on a given day. The floret view was calculated by the number of open florets on a given day. The population schedule was calculated by utilizing the estimated number of florets per day from the model and using that number. Then, all three perspectives were scaled to a proportion for each day that took the total for the day divided by the total for all days, so that they would be comparable.

These three perspectives were compared with the Kolmogorov-Smirnov test to see if they had the same underlying distribution. Then, two measures of synchrony, peak day and middle 50%, were calculated using the three perspectives.

All analyses were performed using R 3.2.3 (R Development Core Team 2015).

RESULTS

Solo vs. Ensemble heads

P-values and confidence intervals from the bootstrap analysis determining if the five variables differ individually with uniqueness of head are shown in table 1, and the five function parameters did not differ between solo and ensemble heads. Based on the MANOVA test, they also do not vary together ($F_{(1,345)} = 0.50$, $p = 0.773$, Wilk's $\Lambda = 0.99$), so both solo and ensemble heads contribute equally to the population's flowering schedule.

Population differences

The results from the MANOVA test and follow up univariate tests are shown in table 2. Based on these results a population of *Echinacea* plants is defined by the site of origin and by the year planted, and this affects the flowering schedule as a whole. Follow-up univariate tests show that the maximum, start date, duration, and lateness differ individually between populations.

Figure 3 shows the flowering schedules of heads divided by 6 populations. Within a population there is variation between heads, but overall trends are reflected in the population flowering schedule.

Comparing Perspectives

According to the Kolmogorov-Smirnov test, the floret perspective and the population schedule have the same underlying distribution ($p=0.998$), whereas the plant perspective is different than both of them ($p<0.001$). This supports the idea that the sine functions are a good model to represent the raw data.

Table 3 shows the peak day synchrony measure according to the three perspectives for 7 populations. Compared to the floret peak day, the schedule perspective matched in 3 populations and was off by an average of 1.5 days on the other 4 populations, whereas the plant peak day never matched, but was only off by an average of 1.6 days. Table 4 shows the middle 50% synchrony measure according to the three perspectives for the 7 populations. Compared to the floret middle 50%, the schedule perspective matched exactly for 4 of the 7 populations, and the three populations that didn't match differed between the perspectives by 1 day in length. The plant middle 50% matched the start date from the floret perspective on 4 of the 7 spans with the other 3 varying by one day, but the end date did not match on any of them and varied by an average of 2.5 days longer. So, the plant perspective is close to the floret perspective on the start and peak, but tends to have a long tail (overestimates the end of the synchrony).

DISCUSSION

Since all of these plants were grown in a common environment, the differences in flowering schedules are not due to environmental factors, but are instead due to genetic differences between the different populations. This has some positive and negative aspects for conservation biology. If you utilize different populations for prairie restoration, the prairie will be genetically diverse which will allow the new population to be more resilient and increase the probability of coming in contact with pollen from non-related *Echinacea*. Since some populations tend to flower earlier, and some flower later, the new population will also flower for longer which will increase the amount of time that food and pollen will be available to the surrounding ecosystem. There are some disadvantages though as well. In Figure 3, you can see that some of the heads in the ERI population have their peak flowering on July 20^{th} or after, when the rest of the populations are done flowering. This puts this population at a disadvantage in the new environment, and it may be outcompeted by the other populations if it is unable to get pollen and create viable seeds.

Within each population, the differences between solo heads and ensemble heads is not significant, so each type of head contributes equally to the overall flowering schedule and no individual head has an advantage or disadvantage just because it is a solo head or ensemble head. Thus, plants with multiple heads have the overall advantage over single-headed plants because they have more opportunities to be pollinated and create viable seeds.

These calculated schedules are good approximations of the raw floret data, but do not match the plant perspective, which is utilized by many people to calculate synchrony. The plant perspective matches closely to the floret data near the beginning of the season, but then overestimates the synchrony potential near the end of the season. Utilizing this new schedule perspective many shine new insights on other experiments that currently use the plant perspective to calculate synchrony measures. Since counting the number of achenes per plant per day is labor intensive and many years and populations do not have this data available, it needs to be investigated if these more accurate flowering schedules can be calculated with different methods, such as duration of flowering and size head, or some other more practical method. It would be beneficial to see if the schedules could be calculated just using the flowering duration and size of the *Echinacea* head.

REFERENCES

Ison, J.L. & Wagenius, S. (2014). Both flowering time and distance to conspecific plants affect reproduction in Echinacea angustifolia, a common prairie perennial. *Journal of Ecology*. **102**, 920-929.

Malo, J.E. (2002). Modelling unimodal flowering phenology with exponential sine equations. *Functional Ecology*. **16**, 413-418.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computer, Vienna.

APPENDIX:

Figure 1. *Histograms of variation in parameters by single headed and multi-headed plants.* Histograms were made from parameters calculated from 145 solo heads and 202 ensemble heads in 2005. The red line represents the mean of each group, and the blue line represents the median.

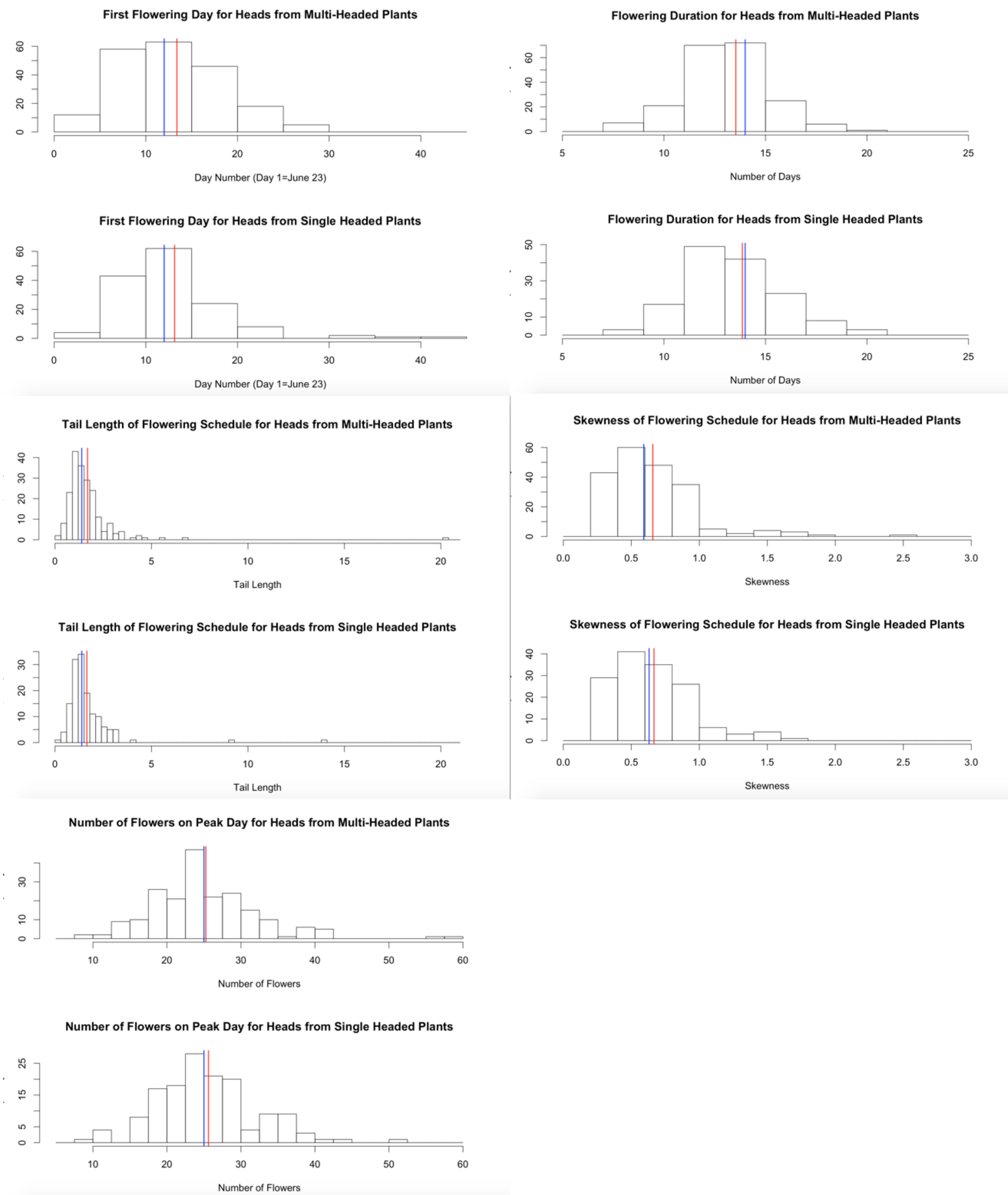
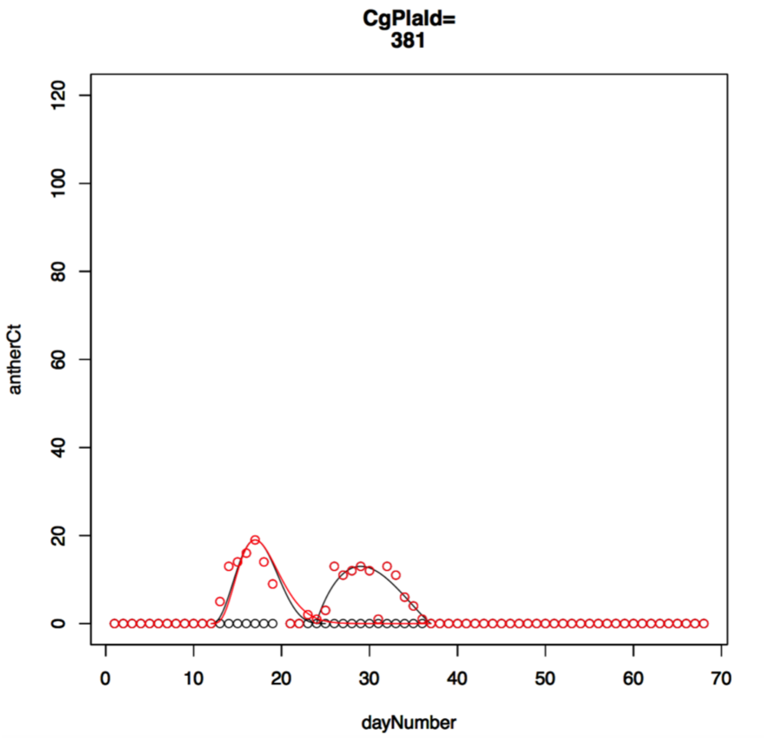


Figure 2. *Graphs of the two dual modal heads.* The black dots represent the raw anther counts for the individual heads with the line representing the sine function fit to the data. The red dots are raw data for the plant as a whole with the red line representing the bad fitting line fit to the flowering schedule of the entire plant.



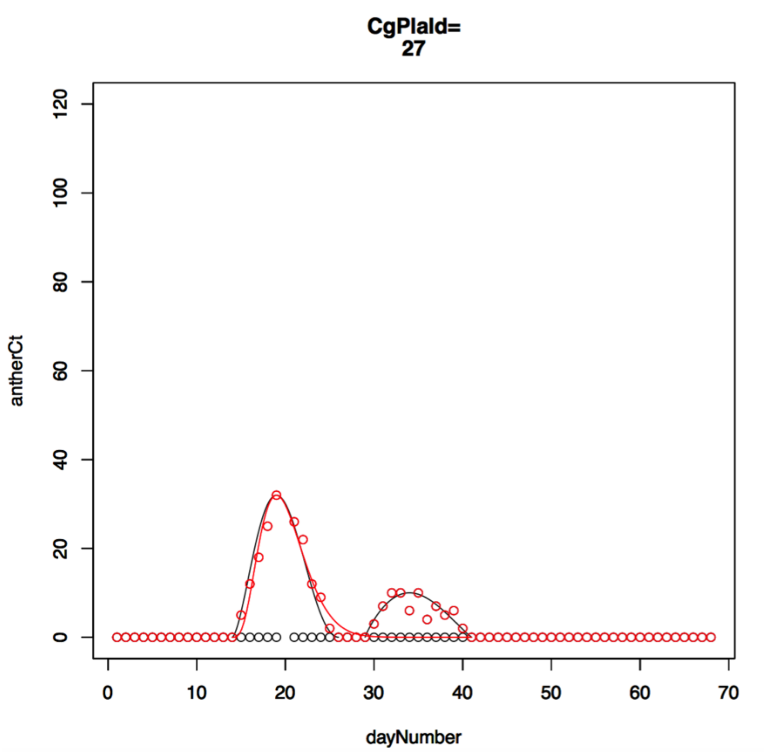


Table 1 *Summary of bootstrap analysis with uniqueness of head to a plant as a predictor*. P-value refers to the getting results as extreme or more extreme than our results if the uniqueness of the head does not affect the variable, and the confidence interval is where 95% of the calculated values fall.

variable	p-value	95% Confidence Interval
a	0.62	24.6-26.1
b	0.67	12.7-13.9
c	0.17	13.4-13.9
d	0.76	0.63-0.69
e	0.80	1.53-1.86

Table 2 *Summary of the MANOVA analysis with year planted and site of origin as a predictor*. The significant predictors are the year planted and the year planted interacting with the site of origin, so a population is defined by both parameters. The column labeled affect shows which variables show significant univariate changes based on the predictor.

Predictor	Wilk's Lambda	P Value	Response
Year Planted	0.85	<0.00001	a, b, c
Site of Origin	0.71	0.2	b, d
Year Planted: Site of Origin	0.82	0.05	c

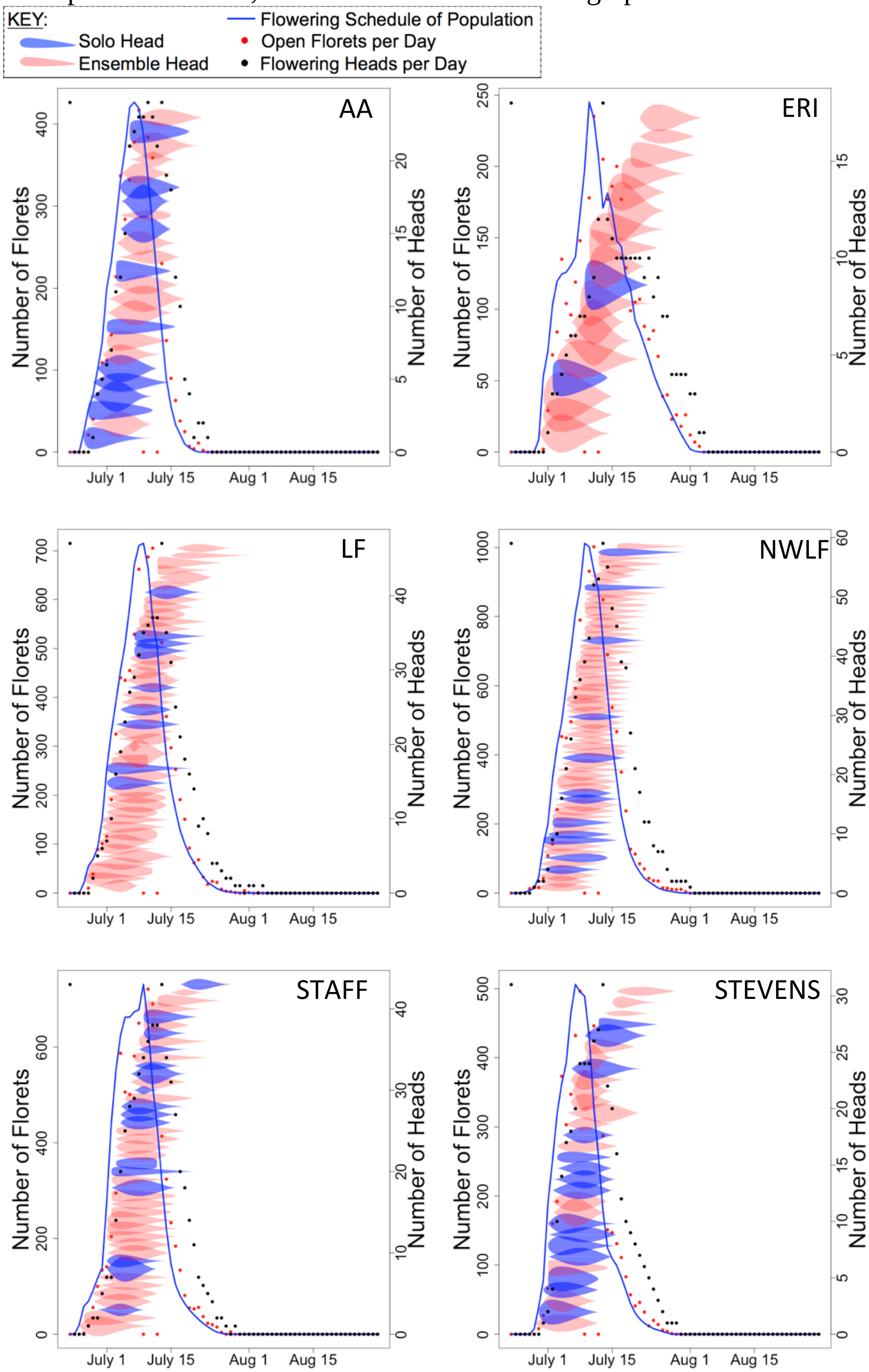
Table 3 *Peak flowering day according to the three different perspectives*. Matching values are highlighted in green.

Population	Plant Peak	Schedule Peak	Floret Peak
AA	20	19	19
cg1	20	18	18
eri	20	16	19
Lf	15	15	16
NWlf	20	18	19
Staff	20	17	18
Stevens	20	15	16

Table 4 *Middle 50% according to the three different perspectives*. The first line for each population is the start date and the second line is the end date for the middle 50%. Matching values are highlighted in green.

Population	Plant	Schedule	Floret
AA	17	16	17
	29	26	26
cg1	18	17	17
	28	24	24
eri	14	14	14
	23	21	21
Lf	12	13	13
	20	19	19
NWlf	16	15	16
	25	22	22
Staff	14	13	13
	22	19	20
Stevens	14	14	14
	22	20	20

Figure 3. *Flowering schedules of heads in 6 populations*. Individual head graphs contain the same overall area as the estimated exponential sine function, with half being represented above and below the set zero line. The red dots represent the raw floret data and the black dots represent the raw plant data, while the rest of the information was calculated. All of these plants were planted in 1996, with site labeled above the graph.



For continued research:
Flog posts: http://echinaceaproject.org/author/rsarette/
Poster: http://echinaceaproject.org/wp-content/uploads/2016/06/rachael_poster.pdf

Data files and reproducible statistical analysis: https://bitbucket.org/swagenius/malocurve. (Initial investigation/data collection done by Stuart with files stored in drop box folder MaloCurveProject.)

Descriptions of files found in repository:
ComparingMaloCurves.R- R script visualizing the variables (a, b, c, d, and e) and seeing if they differ between solo and ensemble heads through bootstrap, nonmetric multi-dimensional scaling, and MANOVA testing. Then paternity and year planted is added into the analysis

Graphing.R- Creating functions to graph all the heads/plants on the same graph, and functions to calculate 3 different perspectives to look at synchrony (plant, floret and estimated value), then comparing the 3 different perspectives.

MultiheadedAnalysis.R- Plotting the graphs for each plant with its heads together, then investigating if the flowering schedules vary between plant populations.

Reference.R- Plotting different values of d and e to see how they affect the shape of the graphs.

malocurve.Rproj- overall project for RStudio

CSV data files (utilized/modified in R files):
MultiheadedPlantOut.csv-from Lou's malo curve program for "MultiheadedAnalysis.R"
data.csv- made in "MultiheadedAnalysis.R" for "Graphing.R"
headInfo.csv- from dropbox for "ComparingMaloCurves.R
multiheadedPlantInfo.csv- made in "MultiheadedAnalysis.R" for Lou's program
plantInfo.csv- from dropbox for "ComparingMaloCurves.R
siteOfOrigin.csv- from "ComparingMaloCurves.R" data originated on Echinacea website
x4.csv- from dropbox for "ComparingMaloCurves.R
x5.csv- from "ComparingMaloCurves.R" for "MultiheadedAnalysis.R"
x6.csv- from "MultiheadedAnalysis.R" for "Graphing.R"

Future directions:
- Does total achene count relate to maximum achenes on a single day?
- Is there a way to estimate the flowering schedule knowing just duration and total achene count?
- Since there are differences in the three statistics does that affect conclusions in other papers, or is it negligible?
- Describe the parameter values for each population (eg. NWLF is the earliest flowering and ERI is the latest) and how do these conclusions differ by perspective?